

Strong Converse Theorems for Classes of Multimessage Multicast Networks: A Rényi Divergence Approach

Silas L. Fong and Vincent Y. F. Tan, *Senior Member, IEEE*

Abstract

This paper establishes that the strong converse holds for some classes of discrete memoryless multimessage multicast networks (DM-MMNs) whose corresponding cut-set bounds are tight, i.e., coincide with the set of achievable rate tuples. The strong converse for these classes of DM-MMNs implies that all sequences of codes with rate tuples belonging to the exterior of the cut-set bound have average error probabilities that necessarily tend to one (and are not simply bounded away from zero). Examples in the classes of DM-MMNs include wireless erasure networks, DM-MMNs consisting of independent discrete memoryless channels (DMCs) as well as single-destination DM-MMNs consisting of independent DMCs with destination feedback. Our elementary proof technique leverages properties of the Rényi divergence.

Index Terms

Strong converse, Multimessage multicast networks, Rényi divergence, Wireless erasure networks

I. INTRODUCTION

This paper considers multimessage multicast networks (MMNs) [1, Chapter 18] in which the destination nodes want to decode the same set of messages transmitted by the source nodes. A well-known outer bound on the capacity region of the discrete memoryless MMN (DM-MMN) is the *cut-set bound*, developed by El Gamal in 1981 [2]. This bound states that for any cut T of the network with nodes indexed by \mathcal{I} , the sum of the achievable rates of messages on one side of the cut is upper bounded by the conditional mutual information of the input variables in T and the output variables in $T^c \triangleq \mathcal{I} \setminus T$ given the input variables in T^c . The DM-MMN is a generalization of the well-studied discrete memoryless relay channel (DM-RC) [3]. It is known that the cut-set bound is not tight in general [4], but it is tight for several classes of DM-MMNs, including the physically degraded DM-RC [3],

Silas L. Fong and Vincent Y. F. Tan are with the Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore (e-mails: {silas_fong, vtan}@nus.edu.sg). Vincent Y. F. Tan is also with the Department of Mathematics, NUS.

The authors are supported by NUS grant R-263-000-A98-750/133.

This paper was presented in part at 2015 Information Theory and Applications (ITA) workshop, San Diego, CA, and 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong.

the semi-deterministic DM-RC [5], the deterministic relay network with no interference [6], the finite-field linear deterministic network [7], [8] and the wireless erasure network [9].

One potential drawback of the cut-set bound is the fact if it is tight, i.e., there exists a matching achievable inner bound, this only implies a *weak converse* for the problem. In other words, it only guarantees that for all rate tuples not belonging to the region prescribed by the cut-set bound, the average error probability in decoding the transmitted messages is bounded away from zero as the block length of any code tends to infinity. In information theory, it is also important to establish *strong converses* as such definitive statements indicate that there is a sharp phase transition between rate tuples that are achievable and those that are not. A strong converse implies that for all codes with rate tuples that are in the exterior of the region prescribed by the fundamental limit, the error probability must necessarily tend to one. The contrapositive of this statement can be stated as follows: All codes whose error probabilities are no larger than $\epsilon \in [0, 1)$ as the block length grows, i.e., ϵ -reliable codes, must have rate tuples belonging to the region prescribed by the fundamental limit (in our case, a looser version of the cut-set bound that is tight for some DM-MMNs). This is clearly a stronger statement than the weak converse which considers codes with vanishing error probabilities.

A. Main Contribution

The main contribution of this work is a self-contained proof of the strong converse for some classes of DM-MMNs in which the cut-set bound is tight. These classes of DM-MMNs include deterministic relay networks with no interference [6], finite-field linear deterministic networks [7], [8] and wireless erasure networks [9]. So for example, for wireless erasure networks studied by Dana, Gowaiker, Palanki, Hassibi and Effros [9], all sequences of codes with rates above the capacity have average error probabilities that necessarily tend to one as the block length grows. The authors of [9] proved using Fano's inequality [10, Section 2.10] that all codes with rates above capacity have average error probabilities that are bounded away from zero. Thus, a consequence of our main result is an important strengthening of the converse in [9, Theorem 2]. In addition, we show, using our main theorem, that the strong converse holds for *DM-MMNs consisting of independent discrete memoryless channels (DMCs) and single-destination DM-MMNs consisting of independent DMCs with destination feedback*. Our main result implies that for the aforementioned DM-MMNs, rate tuples of ϵ -reliable codes where $\epsilon \in [0, 1)$ must belong to the region prescribed by the cut-set bound [2].

The technique that we employ is based on properties of the Rényi divergence [11]–[13]. This is a powerful technique for establishing strong converses in information theory. It has been employed previously to establish strong converses for point-to-point memoryless DMCs with output feedback [11], [14], classical-quantum channels [15] and most recently, entanglement-breaking quantum channels [16]. We were inspired to use the Rényi divergence technique for our strong converse proof because of the similarities of DM-MMNs to channels with full output feedback as shown in the context of sphere-packing bounds on the reliability function for the DM-RC in [17].

B. Related Work

The papers that are most closely related to the present work are the ones by Behboodi and Piantanida who conjectured that the strong converse holds for DM-RCs [18] and general DM multicast networks [19]. Also see Appendix C in the thesis by Behboodi [20]. It appears to the present authors, however, that some steps in the justifications, which are based on the information spectrum method [21], are incomplete. Therefore, we are motivated to provide a strong converse for some (albeit somewhat restrictive) classes of DM-MMNs using a completely different and elementary method—namely, the Rényi divergence approach [11]–[13]. As mentioned by Polyanskiy and Verdú [11], this approach is arguably the simplest method for proving that memoryless channels with feedback satisfy the strong converse and thus, we are inspired to leverage it to prove the strong converse for some classes of DM-MMNs.

C. Paper Outline

This paper is organized as follows. Section II presents the notation used in this paper. Section III provides the problem formulation of the DM-MMNs and presents our main theorem. Section IV introduces the Rényi divergence and discusses its important properties. Section V contains an important lemma concerning *simulating distributions* which is used in the proof of our main theorem. Section VI presents the proof of our main theorem. We also compare and contrast our proof with the proof of the usual cut-set bound which only implies a weak converse. In Section VII, we discuss the above-mentioned classes of DM-MMNs whose cut-set bounds are tight, and we use our main theorem to prove the strong converse for them. We conclude our discussion and suggest avenues for future research in Section VIII. Proofs of the more technical auxiliary results are relegated to the appendices.

II. NOTATION

We use $\Pr\{\mathcal{E}\}$ to represent the probability of an event \mathcal{E} , and we let $\mathbf{1}(\mathcal{E})$ be the characteristic function of \mathcal{E} . We use a capital letter X to denote a random variable with alphabet \mathcal{X} , and use the small letter x to denote a realization of X . We use X^n to denote a random vector $[X_1 \ X_2 \ \dots \ X_n]$, where the components X_k have the same alphabet \mathcal{X} . We let p_X and $p_{Y|X}$ denote the probability mass distribution of X and the conditional probability mass distribution of Y given X respectively for any discrete random variables X and Y . For any mapping g whose domain includes \mathcal{X} , we let $p_{g(X)}$ denote the probability mass distribution of $g(X)$ when X is distributed according to p_X . We let $p_X(x) \triangleq \Pr\{X = x\}$ and $p_{Y|X}(y|x) \triangleq \Pr\{Y = y|X = x\}$ be the evaluations of p_X and $p_{Y|X}$ respectively at $X = x$ and $Y = y$. We let $p_{XY|X}$ denote the joint distribution of (X, Y) , i.e., $p_{XY|X}(x, y) = p_X(x)p_{Y|X}(y|x)$ for all x and y . If X and Y are independent, their joint distribution is simply $p_X p_Y$. For simplicity, we drop the subscript of a notation if there is no ambiguity. We will take all logarithms to base 2, and we will use the convention that $0 \log 0 = 0$ and $0 \log \frac{0}{0} = 0$ throughout this paper. For any discrete random variable (X, Y, Z) distributed according to $p_{X,Y,Z}$, we let $H_{p_{X,Z}}(X|Z)$ and $I_{p_{X,Y,Z}}(X; Y|Z)$ be the entropy of X given Z and mutual information between X and Y given Z respectively. The \mathcal{L}_1 -distance between two distributions p_X and q_X on the same discrete alphabet \mathcal{X} , denoted by $\|p_X - q_X\|_{\mathcal{L}_1}$, is defined as $\|p_X - q_X\|_{\mathcal{L}_1} \triangleq \sum_{x \in \mathcal{X}} |p_X(x) - q_X(x)|$. If X ,

Y and Z are distributed according to $p_{X,Y,Z}$ and they form a Markov chain, we write $(X \rightarrow Y \rightarrow Z)_{p_{X,Y,Z}}$ or more simply, $(X \rightarrow Y \rightarrow Z)_p$.

III. PROBLEM FORMULATION AND MAIN RESULT

We consider a DM-MMN that consists of N nodes. Let

$$\mathcal{I} \triangleq \{1, 2, \dots, N\}$$

be the index set of the nodes, and let $\mathcal{S} \subseteq \mathcal{I}$ and $\mathcal{D} \subseteq \mathcal{I}$ be the sets of sources and destinations respectively. We call $(\mathcal{S}, \mathcal{D})$ the *multicast demand* on the network. The sources in \mathcal{S} transmit information to the destinations in \mathcal{D} in n time slots (channel uses) as follows. Node i transmits message

$$W_i \in \{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$$

for each $i \in \mathcal{S}$ and node j , for each $j \in \mathcal{D}$, wants to decode $\{W_i : i \in \mathcal{S}\}$, where R_i denotes the rate of message W_i . We assume that each message W_i is uniformly distributed over $\{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$ and all the messages are independent. For each time slot $k \in \{1, 2, \dots, n\}$ and each $i \in \mathcal{I}$, node i transmits $X_{i,k} \in \mathcal{X}_i$, a function of (W_i, Y_i^{k-1}) , and receives, from the output of a channel, $Y_{i,k} \in \mathcal{Y}_i$ where \mathcal{X}_i and \mathcal{Y}_i are some alphabets that possibly depend on i . After n time slots, node j declares $\hat{W}_{i,j}$ to be the transmitted W_i based on (W_j, Y_j^n) for each $(i, j) \in \mathcal{S} \times \mathcal{D}$.

To simplify notation, we use the following conventions for each $T \subseteq \mathcal{I}$: For any random tuple

$$(X_1, X_2, \dots, X_N) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N,$$

we let

$$X_T \triangleq (X_i : i \in T)$$

be a subtuple of (X_1, X_2, \dots, X_N) . Similarly, for any $k \in \{1, 2, \dots, n\}$ and any random tuple

$$(X_{1,k}, X_{2,k}, \dots, X_{N,k}) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N,$$

we let

$$X_{T,k} \triangleq (X_{i,k} : i \in T)$$

be a subtuple of $(X_{1,k}, X_{2,k}, \dots, X_{N,k})$. For any N^2 -dimensional random tuple $(\hat{W}_{1,1}, \hat{W}_{1,2}, \dots, \hat{W}_{N,N})$, we let

$$\hat{W}_{T \times T^c} \triangleq (\hat{W}_{i,j} : (i, j) \in T \times T^c)$$

be a subtuple of $(\hat{W}_{1,1}, \hat{W}_{1,2}, \dots, \hat{W}_{N,N})$.

The following six definitions formally define a DM-MMN and its capacity region.

Definition 1: A discrete network consists of N finite input sets $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_N$, N finite output sets $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_N$

and a conditional distribution $q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}$. The discrete network is denoted by $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$.

Definition 2: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a discrete network, and let $(\mathcal{S}, \mathcal{D})$ be the multicast demand on the network. An $(n, R_{\mathcal{I}})$ -code, where $R_{\mathcal{I}}$ denotes the tuple of code rates (R_1, R_2, \dots, R_N) , for n uses of the network consists of the following:

- 1) A message set

$$\mathcal{W}_i = \{1, 2, \dots, \lceil 2^{nR_i} \rceil\}$$

at node i for each $i \in \mathcal{I}$, where $R_i = 0$ for each $i \in \mathcal{S}^c$. Message W_i is uniform on \mathcal{W}_i .

- 2) An encoding function

$$\phi_{i,k} : \mathcal{W}_i \times \mathcal{Y}_i^{k-1} \rightarrow \mathcal{X}_i$$

for each $i \in \mathcal{I}$ and each $k \in \{1, 2, \dots, n\}$, where $\phi_{i,k}$ is the encoding function at node i in the k^{th} time slot such that

$$X_{i,k} = \phi_{i,k}(W_i, Y_i^{k-1}).$$

- 3) A decoding function

$$\psi_{i,j} : \mathcal{W}_j \times \mathcal{Y}_j^n \rightarrow \mathcal{W}_i$$

for each $(i, j) \in \mathcal{S} \times \mathcal{D}$, where $\psi_{i,j}$ is the decoding function for message W_i at node j such that

$$\hat{W}_{i,j} = \psi_{i,j}(W_j, Y_j^n).$$

Since the encoder $\phi_{i,k}$ can depend on the “feedback signal” Y_i^{k-1} , we are allowing full output feedback for each of the transmitting nodes; cf. Section VII-C. In addition, the definition of $\phi_{i,k}$ allows every node to process information in a causal way with a delay of one unit.

Definition 3: A discrete network $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ with multicast demand $(\mathcal{S}, \mathcal{D})$, when used multiple times, is called a *discrete memoryless multimessage multicast network (DM-MMN)* if the following holds for any $(n, R_{\mathcal{I}})$ -code:

For all $T \subseteq \mathcal{I}$, we define $q_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c}|x_{\mathcal{I}})$, the marginal distribution of channel $q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}(y_{\mathcal{I}}|x_{\mathcal{I}})$, as follows:

$$q_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c}|x_{\mathcal{I}}) \triangleq \sum_{y_T \in \mathcal{Y}_T} q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}(y_{\mathcal{I}}|x_{\mathcal{I}})$$

for all $x_{\mathcal{I}} \in \mathcal{X}_{\mathcal{I}}$ and $y_{T^c} \in \mathcal{Y}_{T^c}$. Let $U^{k-1} \triangleq (W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}, Y_{\mathcal{I}}^{k-1})$ be the collection of random variables that are generated before the k^{th} time slot. Then, for each $k \in \{1, 2, \dots, n\}$ and each $T \subseteq \mathcal{I}$,

$$\Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_{\mathcal{I},k}, Y_{T^c,k} = y_{T^c,k}\} = \Pr\{U^{k-1} = u^{k-1}, X_{\mathcal{I},k} = x_{\mathcal{I},k}\} q_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c,k}|x_{\mathcal{I},k}) \quad (1)$$

for all $u^{k-1} \in \mathcal{U}^{k-1}$, $x_{\mathcal{I},k} \in \mathcal{X}_{\mathcal{I}}$ and $y_{T^c,k} \in \mathcal{Y}_{T^c}$.

Definition 4: For an $(n, R_{\mathcal{I}})$ -code defined on the DM-MMN with multicast demand $(\mathcal{S}, \mathcal{D})$, the *average proba-*

bility of decoding error is defined as

$$\Pr \left\{ \bigcup_{j \in \mathcal{D}} \bigcup_{i \in \mathcal{S}} \{\hat{W}_{i,j} \neq W_i\} \right\}.$$

We call an $(n, R_{\mathcal{I}})$ -code with average probability of decoding error not exceeding ϵ_n an $(n, R_{\mathcal{I}}, \epsilon_n)$ -code.

Definition 5: A rate tuple $R_{\mathcal{I}}$ is ϵ -achievable for the DM-MMN with multicast demand $(\mathcal{S}, \mathcal{D})$ if there exists a sequence of $(n, R_{\mathcal{I}}, \epsilon_n)$ -codes for the DM-MMN such that

$$\limsup_{n \rightarrow \infty} \epsilon_n \leq \epsilon.$$

Definition 6: The ϵ -capacity region (for $\epsilon \in [0, 1)$) of the DM-MMN with multicast demand $(\mathcal{S}, \mathcal{D})$, denoted by \mathcal{C}_ϵ , is the set consisting of all ϵ -achievable rate tuples $R_{\mathcal{I}}$ with $R_i = 0$ for all $i \in \mathcal{S}^c$. The capacity region is defined to be the 0-capacity region \mathcal{C}_0 .

The following theorem is the main result in this paper.

Theorem 1: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a DM-MMN with multicast demand $(\mathcal{S}, \mathcal{D})$. Define

$$\mathcal{R}_{\text{out}} \triangleq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \bigcup_{p_{X_{\mathcal{I}}}} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (2)$$

Then for each $\epsilon \in [0, 1)$,

$$\mathcal{C}_\epsilon \subseteq \mathcal{R}_{\text{out}}. \quad (3)$$

We now make a couple of remarks concerning Theorem 1.

First, define the usual cut-set bound [1, Theorem 18.1]

$$\mathcal{R}_{\text{cut-set}} \triangleq \bigcup_{p_{X_{\mathcal{I}}}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (4)$$

It is well known that $\mathcal{R}_{\text{cut-set}}$ is an outer bound on the capacity region, i.e., that

$$\mathcal{C}_0 \subseteq \mathcal{R}_{\text{cut-set}}. \quad (5)$$

Note that \mathcal{R}_{out} is similar to $\mathcal{R}_{\text{cut-set}}$ except that the union and the intersection operations are interchanged. Consequently, \mathcal{R}_{out} is potentially looser (larger) than the $\mathcal{R}_{\text{cut-set}}$. This discrepancy is briefly explained as follows: The proof of Theorem 1 (i.e., the bound in (3)) leverages the properties of the Rényi divergence, while the proof of the cut-set bound (i.e., the bound in (5)) is based on Fano's inequality [1, Theorem 18.1]. For both proofs, the first step is to fix an achievable rate tuple $R_{\mathcal{I}}$ and a sequence of $(n, R_{\mathcal{I}})$ -codes. Next a cut $T \subseteq \mathcal{I}$ that satisfies $T^c \cap \mathcal{D} \neq \emptyset$ is also fixed. In both proofs, we eventually arrive at the bound

$$\sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}}}^{(T)}(X_T; Y_{T^c} | X_{T^c})$$

for some $p_{X_{\mathcal{I}}}^{(T)}$, which implies that

$$\mathcal{C}_\epsilon \subseteq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \bigcup_{p_{X_{\mathcal{I}}}^{(T)}} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}}^{(T)} q_{Y_{T^c} | X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (6)$$

However, the proofs of bounds (3) and (5) yield (6) under different assumptions on the asymptotic behavior of the average error probability ϵ . For the proof of the cut-set bound (5), it is assumed that $\epsilon = 0$ and hence using Fano's inequality combined with properties of the relative entropy and the conditional mutual information such as the chain rule are sufficient for proving (6). Using Fano's inequality, $p_{X_{\mathcal{I}}}^{(T)}$ can be shown to be the limit of the sequence of empirical input distributions induced by the sequence of codes (if the limit does not exist, we can always consider a convergent subsequence instead and the following arguments go through in a similar way). In other words,

$$p_{X_{\mathcal{I}}}^{(T)}(x_{\mathcal{I}}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{X_{\mathcal{I}},k}(x_{\mathcal{I}}).$$

This implies that $p_{X_{\mathcal{I}}}^{(T)}$ does *not* depend on T and hence the union and the intersection operations in (6) can be interchanged, resulting in an improved bound (5). In contrast, for the proof of our bound (3), it is assumed that $\epsilon \in [0, 1)$ and hence we need to use properties of the Rényi divergence to prove (6). Since $p_{X_{\mathcal{I}}}^{(T)}$ *does* depend on T in general for the proof involving the Rényi divergence, the union and the intersection operations in (6) cannot be interchanged in general, which prevents us from further strengthening the bound in (3). In Section VI-G, we further elaborate on the similarities of and differences between the proofs of our bound (3) and the cut-set bound (5).

Second, although \mathcal{R}_{out} is potentially looser than the cut-set bound, it can be shown that $\mathcal{R}_{\text{out}} \subseteq \mathcal{C}_\epsilon$ for some classes of networks including the deterministic relay networks with no interference [6], the finite-field linear deterministic networks [7], [8] and the wireless erasure networks [9] (discussed in Section VII-A), the class of DM-MMNs consisting of independent DMCs (discussed in Section VII-B) and the class of single-destination DM-MMNs consisting of independent DMCs with destination feedback (discussed in Section VII-C). Therefore, Theorem 1 implies the strong converses for these networks.

We briefly outline the content in the sections to follow: The proof of Theorem 1 leverages properties of the Rényi divergence, which we discuss in Section IV. In Section V, we construct so-called *simulating distributions*, which form an important part of the proof of Theorem 1. The details of the proof of Theorem 1 are provided in Section VI. Readers who are only interested in the application of Theorem 1 to specific channel models may proceed directly to Section VII.

IV. PROPERTIES OF THE RÉNYI DIVERGENCE

The following definitions of (conditional) relative entropy and (conditional) Rényi divergence are standard [11]–[13].

Definition 7: Let p_X and q_X be two probability distributions on \mathcal{X} , and let r_Z be a probability distribution on

\mathcal{Z} . Let

$$D(p_X \| q_X) \triangleq \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}$$

be the *relative entropy* between p_X and q_X , and let

$$D(p_{X|Z} \| q_{X|Z} | r_Z) \triangleq \sum_{z \in \mathcal{Z}} r_Z(z) D(p_{X|Z=z} \| q_{X|Z=z})$$

be the *conditional relative entropy* between $p_{X|Z}$ and $q_{X|Z}$ conditioned on r_Z . Then, the *Rényi divergence with parameter* $\lambda \in [1, \infty)$ between p_X and q_X , denoted by $D_\lambda(p_X \| q_X)$, is defined as follows:

$$D_\lambda(p_X \| q_X) \triangleq \begin{cases} \frac{1}{\lambda-1} \log \sum_{x \in \mathcal{X}} \frac{(p_X(x))^\lambda}{(q_X(x))^{\lambda-1}} & \text{if } \lambda > 1, \\ D(p_X \| q_X) & \text{if } \lambda = 1. \end{cases}$$

In addition, the *conditional Rényi divergence with parameter* $\lambda \in [1, \infty)$ between $p_{X|Z}$ and $q_{X|Z}$ given r_Z , denoted by $D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z)$, is defined as follows:

$$D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z) \triangleq \begin{cases} \frac{1}{\lambda-1} \log \sum_{z \in \mathcal{Z}} r_Z(z) \sum_{x \in \mathcal{X}} \frac{(p_{X|Z}(x|z))^\lambda}{(q_{X|Z}(x|z))^{\lambda-1}} & \text{if } \lambda > 1, \\ D(p_{X|Z} \| q_{X|Z} | r_Z) & \text{if } \lambda = 1. \end{cases}$$

Note that for $\lambda > 1$, $D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z)$ can be expressed in terms of the unconditional Rényi divergence as

$$D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z) = \frac{1}{\lambda-1} \log \sum_{z \in \mathcal{Z}} r_Z(z) 2^{(\lambda-1) D_\lambda(p_{X|Z=z} \| q_{X|Z=z})}.$$

We summarize two important properties of $D_\lambda(p_{X|Z} \| q_{X|Z} | r_Z)$ in the following theorem, whose proof can be found in [22, Theorems 5 and 9].

Theorem 2: For any $\lambda \in [1, \infty)$, the following statements hold for any two conditional probability distributions $p_{X,Y|Z}$, $q_{X,Y|Z}$ and any probability distribution r_Z :

1. (Continuity) $D_\lambda(p_{X|Z} \| q_{Y|Z} | r_Z)$ is continuous in λ .
2. (Data processing inequality (DPI)) $D_\lambda(p_X \| q_X) \geq D_\lambda(p_{g(X)} \| q_{g(X)})$ for any function g with domain \mathcal{X} . In particular, $D_\lambda(p_{X,Y} \| q_{X,Y}) \geq D_\lambda(p_X \| q_X)$.

Most converse theorems use Fano's inequality [23, Section 2.10] to obtain a lower bound on the error probability. However, this can only lead to weak converse results. The following proposition, analogous to Fano's inequality, enables us to prove strong converse results by providing a better lower bound on the error probability. Essentially, we have the freedom to choose any $\lambda \in (1, \infty)$ in the bound in (8) below.

Proposition 1: Let $p_{U,V}$ be a probability distribution defined on $\mathcal{W} \times \mathcal{W}$ for some \mathcal{W} , and let p_U be the marginal distribution of $p_{U,V}$. In addition, let q_V be a distribution defined on \mathcal{W} . Suppose p_U is the uniform distribution, and let

$$\alpha = \Pr\{U \neq V\} \tag{7}$$

be a real number in $[0, 1)$. Then for each $\lambda \in (1, \infty)$,

$$D_\lambda(p_{U,V} \| p_U q_V) \geq \log |\mathcal{W}| + \lambda(\lambda - 1)^{-1} \log(1 - \alpha). \quad (8)$$

Proof: Fix a $\lambda \in (1, \infty)$ and let $s_{U,V} \triangleq p_U q_V$. Consider the following chain of inequalities:

$$\begin{aligned} D_\lambda(p_{U,V} \| s_{U,V}) &\stackrel{(a)}{\geq} D_\lambda(p_{\mathbf{1}_{\{U=V\}}} \| s_{\mathbf{1}_{\{U=V\}}}) \\ &\stackrel{(b)}{=} \frac{1}{\lambda - 1} \log \left(|\mathcal{W}|^{\lambda-1} (1 - \alpha)^\lambda + \left(\frac{|\mathcal{W}|}{|\mathcal{W}| - 1} \right)^{\lambda-1} \alpha^\lambda \right) \\ &\stackrel{(c)}{\geq} \log |\mathcal{W}| + \lambda(\lambda - 1)^{-1} \log(1 - \alpha), \end{aligned}$$

where

(a) follows from the DPI in Theorem 2;

(b) follows from Definition 7 and the facts that

$$\begin{aligned} \sum_{u,v} p_{U,V}(u, v) \mathbf{1}(\{u = v\}) &= 1 - \alpha \quad (\text{cf. (7)}) \quad \text{and} \\ \sum_{u,v} s_{U,V}(u, v) \mathbf{1}(\{u = v\}) &= \frac{1}{|\mathcal{W}|}; \end{aligned}$$

(c) follows from the fact that $\left(\frac{|\mathcal{W}|}{|\mathcal{W}| - 1} \right)^{\lambda-1} \alpha^\lambda \geq 0$.

This completes the proof. ■

The following proposition enables us to approximate the conditional Rényi divergence D_λ by the conditional relative entropy $D_1 = D$. Since the proof for the following proposition is straightforward but involves some tedious algebra, we defer it to Appendix A.

Proposition 2: Let $\lambda \in [1, 5/4]$ be a real number, and let $p_{X,Y,Z}$ be a probability distribution defined on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Then,

$$D_\lambda(p_{X,Y|Z} \| p_{X|Z} p_{Y|Z} | p_Z) \leq D(p_{X,Y|Z} \| p_{X|Z} p_{Y|Z} | p_Z) + 8(\lambda - 1)(|\mathcal{X}||\mathcal{Y}|)^5. \quad (9)$$

We made no attempt to optimize the remainder term $8(|\mathcal{X}||\mathcal{Y}|)^5$ as the important part of the statement is that this remainder term is uniform in $p_{X,Y,Z}$ on a sufficiently small interval to the right of $\lambda = 1$. In fact, it only depends on the product $|\mathcal{X}||\mathcal{Y}|$.

V. SIMULATING DISTRIBUTION

Proposition 1 provides a lower bound for the error probability, and the lower bound holds for all q_V . Therefore, we are motivated to choose a *simulating distribution* q_V so that the left hand side of (8) can be simplified. Before describing the simulating distribution, we state the following proposition which facilitates to characterize an important property of Markov chains.

Proposition 3: Suppose there exist two probability distributions $r_{X,Y}$ and $q_{Z|Y}$ such that

$$p_{X,Y,Z}(x, y, z) = r_{X,Y}(x, y)q_{Z|Y}(z|y) \quad (10)$$

for all x, y and z whenever $p_Y(y) > 0$. Then

$$(X \rightarrow Y \rightarrow Z)_{p_{X,Y,Z}} \quad (11)$$

forms a Markov chain. In addition,

$$p_{Z|Y} = q_{Z|Y}. \quad (12)$$

Proof: The proof of (11) is contained [24, Proposition 2.5]. It remains to show (12). Summing x and then z on both sides of (10), we have $p_{Y,Z}(y, z) = r_Y(y)q_{Z|Y}(z|y)$ and $p_Y(y) = r_Y(y)$ for all x, y and z whenever $p_Y(y) > 0$, which implies (12). ■

The construction of the simulating distribution is contained in the following lemma. Before stating lemma, we make the following definitions: Given an $(n, R_{\mathcal{I}}, \epsilon_n)$ -code, we let $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ be the probability distribution induced by the code according to Definitions 2 and 3. In the following, we drop the subscripts of the probability distributions to simplify notation. For each $T \subseteq \mathcal{I}$ and each $\lambda \in [1, \infty)$, recalling that $q_{Y_{T^c}|X_{\mathcal{I}}}$ denotes the channel of the DM-MMN defined in Definition 3, we define $s_{X_{\mathcal{I},1}, Y_{T^c,1}}^{(\lambda, T)} \triangleq p_{X_{\mathcal{I},1}} q_{Y_{T^c}|X_{\mathcal{I}}}$. Then, we define $s_{X_{\mathcal{I},k}}^{(\lambda, T)}$ and $s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda, T)}$ for $k = 2, \dots, n$ based on $\{s_{X_{\mathcal{I},\ell}, Y_{T^c,\ell}}^{(\lambda, T)}\}_{\ell=1}^{k-1}$ as follows: For all $x_{\mathcal{I},k} \in \mathcal{X}_{\mathcal{I}}$ and $y_{T^c,k} \in \mathcal{Y}_{T^c}$, let

$$s_{X_{\mathcal{I},k}}^{(\lambda, T)}(x_{\mathcal{I},k}) \triangleq \frac{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} p(x_{\mathcal{I},k} | x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}) \prod_{\ell=1}^{k-1} \left(p(x_{\mathcal{I},\ell} | x_{\mathcal{I}}^{\ell-1}, y_{T^c}^{\ell-1}) \frac{(q(y_{T^c,\ell} | x_{\mathcal{I},\ell}))^\lambda}{(s_{X_{\mathcal{I},\ell}, Y_{T^c,\ell}}^{(\lambda, T)}(y_{T^c,\ell} | x_{\mathcal{I},\ell}))^{\lambda-1}} \right)}{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} \prod_{\ell=1}^{k-1} \left(p(x_{\mathcal{I},\ell} | x_{\mathcal{I}}^{\ell-1}, y_{T^c}^{\ell-1}) \frac{(q(y_{T^c,\ell} | x_{\mathcal{I},\ell}))^\lambda}{(s_{X_{\mathcal{I},\ell}, Y_{T^c,\ell}}^{(\lambda, T)}(y_{T^c,\ell} | x_{\mathcal{I},\ell}))^{\lambda-1}} \right)} \quad (13)$$

and

$$s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda, T)}(x_{\mathcal{I},k}, y_{T^c,k}) \triangleq s_{X_{\mathcal{I},k}}^{(\lambda, T)}(x_{\mathcal{I},k}) q_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c,k} | x_{\mathcal{I},k}). \quad (14)$$

It can be verified by using (1), (13) and (14) that $s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(1, T)} = p_{X_{\mathcal{I},k}, Y_{T^c,k}}$, and hence $s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda, T)}$ can be viewed as a tilted version of $p_{X_{\mathcal{I},k}, Y_{T^c,k}}$. More specifically, we can see from (13) that $s_{X_{\mathcal{I},k}}^{(\lambda, T)}$ can be viewed as a weighted version of $p_{X_{\mathcal{I},k} | X_{\mathcal{I}}^{k-1}, Y_{T^c}^{k-1}}$ where the weighting distribution is a tilting of $\prod_{\ell=1}^{k-1} (p_{X_{\mathcal{I},\ell} | X_{\mathcal{I}}^{\ell-1}, Y_{T^c}^{\ell-1}} q_{Y_{T^c}|X_{\mathcal{I}}})$ towards $\prod_{\ell=1}^{k-1} (p_{X_{\mathcal{I},\ell} | X_{\mathcal{I}}^{\ell-1}, Y_{T^c}^{\ell-1}} s_{Y_{T^c,\ell} | X_{\mathcal{I},\ell}}^{(\lambda, T)})$.

Lemma 4: Given an $(n, R_{\mathcal{I}}, \epsilon_n)$ -code for the DM-MMN, let $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ be the probability distribution induced by the code according to Definitions 2 and 3. Let T be an arbitrary subset of \mathcal{I} and fix an arbitrary $\lambda \in [1, \infty)$. Then there exists a probability distribution $s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ that satisfies the following properties:

- (i) $s_{W_{\mathcal{I}}} = p_{W_{\mathcal{I}}}$.
- (ii) $s_{\hat{W}_{T \times T^c} | W_{T^c}, Y_{T^c}^n} = p_{\hat{W}_{T \times T^c} | W_{T^c}, Y_{T^c}^n}$.
- (iii) For each $k \in \{1, 2, \dots, n\}$, $((W_{\mathcal{I}}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}) \rightarrow X_{T^c,k} \rightarrow Y_{T^c,k})_s$ forms a Markov chain.
- (iv) For each $k \in \{1, 2, \dots, n\}$, $s_{Y_{T^c,k} | X_{T^c,k}} = s_{Y_{T^c,k} | X_{T^c,k}}^{(\lambda, T)}$, where $s_{Y_{T^c,k} | X_{T^c,k}}^{(\lambda, T)}$ is induced by the joint distribution

in (14).

(v) For each $k \in \{1, 2, \dots, n\}$, $p_{X_{T^c,k}|W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}, Y_{T^c}^{k-1}} = s_{X_{T^c,k}|W_{T^c}, X_{T^c}^{k-1} Y_{T^c}^{k-1}}$.

We call $s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ a λ -simulating distribution of $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ neglecting T because $s_{X_{\mathcal{I},k}, Y_{T^c,k}}$ represents a “ λ -tilting” of $p_{X_{\mathcal{I},k}, Y_{T^c,k}}$ through Property (iv) and captures all the important properties of $(X_{T^c}^n, Y_{T^c}^n)$ when $(X_{T^c}^n, Y_{T^c}^n)$ is generated according to the given code distribution $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$.

Proof: We prove the lemma by first constructing a distribution of $(W_{\mathcal{I}}, X_{T^c}^n, Y_{T^c}^n)$ denoted by r . Subsequently, we use r as a building block to construct a distribution of $(W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}})$. Define

$$r_{W_{\mathcal{I}}, X_{T^c}^1, Y_{T^c}^1} \triangleq p_{W_{\mathcal{I}}} p_{X_{T^c,1}|W_{T^c}} s_{Y_{T^c,1}|X_{T^c,1}}^{(\lambda, T)}. \quad (15)$$

Recursively construct

$$r_{W_{\mathcal{I}}, X_{T^c}^k, Y_{T^c}^k} \triangleq r_{W_{\mathcal{I}}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} p_{X_{T^c,k}|W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} s_{Y_{T^c,k}|X_{T^c,k}}^{(\lambda, T)} \quad (16)$$

for each $k = 2, 3, \dots, n$, where $s_{Y_{T^c,k}|X_{T^c,k}}^{(\lambda, T)}$ is as defined in (14). Applying (16) recursively from $k = 2$ to $k = n$ and using (15), we have

$$r_{W_{\mathcal{I}}, X_{T^c}^n, Y_{T^c}^n} = p_{W_{\mathcal{I}}} \prod_{k=1}^n \left(p_{X_{T^c,k}|W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} s_{Y_{T^c,k}|X_{T^c,k}}^{(\lambda, T)} \right). \quad (17)$$

After defining r through (15), (16) and (17), we are now ready to define s as follows:

$$s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}} \triangleq p_{X_{T^c}^n, Y_{T^c}^n, \hat{W}_{(T \times T^c)^c}} r_{W_{\mathcal{I}}, X_{T^c}^n, Y_{T^c}^n} p_{\hat{W}_{T \times T^c}|W_{T^c}, Y_{T^c}^n}. \quad (18)$$

In the rest of the proof, we want to show that $s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ satisfies Properties (i), (ii), (iii), (iv) and (v).

Since

$$\begin{aligned} & \sum_{x_{\mathcal{I}}^n, y_{\mathcal{I}}^n, \hat{w}_{\mathcal{I} \times \mathcal{I}}} s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}(w_{\mathcal{I}}, x_{\mathcal{I}}^n, y_{\mathcal{I}}^n, \hat{w}_{\mathcal{I} \times \mathcal{I}}) \\ & \stackrel{(18)}{=} \sum_{x_{T^c}^n, y_{T^c}^n} r_{W_{\mathcal{I}}, X_{T^c}^n, Y_{T^c}^n}(w_{\mathcal{I}}, x_{T^c}^n, y_{T^c}^n) \\ & \stackrel{(17)}{=} p_{W_{\mathcal{I}}}(w_{\mathcal{I}}) \end{aligned}$$

for all $w_{\mathcal{I}}$, it follows that Property (i) holds.

In order to prove Property (ii), we write

$$s_{W_{T^c}, Y_{T^c}^n, \hat{W}_{T \times T^c}} \stackrel{(18)}{=} r_{W_{T^c}, Y_{T^c}^n} p_{\hat{W}_{T \times T^c}|W_{T^c}, Y_{T^c}^n}$$

which implies from Proposition 3 that $s_{\hat{W}_{T \times T^c}|W_{T^c}, Y_{T^c}^n} = p_{\hat{W}_{T \times T^c}|W_{T^c}, Y_{T^c}^n}$.

In order to prove Properties (iii), (iv) and (v), we write for each $k \in \{1, 2, \dots, n\}$

$$\begin{aligned} s_{W_{\mathcal{I}}, X_{T^c}^k, Y_{T^c}^k} &\stackrel{(18)}{=} r_{W_{\mathcal{I}}, X_{T^c}^k, Y_{T^c}^k} \\ &\stackrel{(a)}{=} p_{W_{\mathcal{I}}} \prod_{m=1}^k \left(p_{X_{T^c, m} | W_{T^c}, X_{T^c}^{m-1}, Y_{T^c}^{m-1}} s_{Y_{T^c, m} | X_{T^c, m}}^{(\lambda, T)} \right), \end{aligned} \quad (19)$$

where (a) follows from marginalizing (17). It then follows from (19) and Proposition 3 that for each $k \in \{1, 2, \dots, n\}$,

$$((W_{\mathcal{I}}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}) \rightarrow X_{T^c, k} \rightarrow Y_{T^c, k})_s \quad (20)$$

forms a Markov chain and

$$s_{Y_{T^c, k} | X_{T^c, k}} = s_{Y_{T^c, k} | X_{T^c, k}}^{(\lambda, T)}. \quad (21)$$

Properties (iii) and (iv) follow from (20) and (21) respectively. In addition, for each $k \in \{1, 2, \dots, n\}$,

$$s_{W_{\mathcal{I}}, X_{T^c}^k, Y_{T^c}^{k-1}} \stackrel{(19)}{=} p_{W_{\mathcal{I}}} p_{X_{T^c, k} | W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} \prod_{m=1}^{k-1} \left(p_{X_{T^c, m} | W_{T^c}, X_{T^c}^{m-1}, Y_{T^c}^{m-1}} s_{Y_{T^c, m} | X_{T^c, m}}^{(\lambda, T)} \right). \quad (22)$$

Then, for each $k \in \{1, 2, \dots, n\}$,

$$\begin{aligned} p_{X_{T^c, k} | W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}, Y_{T^c}^{k-1}} &\stackrel{(a)}{=} p_{X_{T^c, k} | W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} \\ &\stackrel{(b)}{=} s_{X_{T^c, k} | W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1}} \end{aligned} \quad (23)$$

where

(a) follows from the fact that $((W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}) \rightarrow (W_{T^c}, Y_{T^c}^{k-1}) \rightarrow X_{T^c, k})_p$ forms a Markov chain (cf. Definition 2).

(b) follows from (22) and Proposition 3.

Property (v) follows from (23). ■

VI. PROOF OF THEOREM 1

We partition the proof into several subsections for the sake of clarity and readability. In the final subsection (Section VI-G), we compare and contrast the proof of Theorem 1 with the proof of the usual cut-set bound which only implies a weak converse.

A. Lower Bounding the Error Probability in Terms of the Rényi Divergence

Fix an $\epsilon \in [0, 1)$ and let $R_{\mathcal{I}}$ be an ϵ -achievable rate tuple for the DM-MMN. By Definitions 5 and 6, there exists a number $\bar{\epsilon} \in [0, 1)$ and a sequence of $(n, R_{\mathcal{I}}, \epsilon_n)$ -codes on the DM-MMN such that for all sufficiently large n ,

$$\epsilon_n \leq \bar{\epsilon}. \quad (24)$$

Fix a sufficiently large n such that (24) holds, and let $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ be the probability distribution induced by the $(n, R_{\mathcal{I}}, \epsilon_n)$ -code on the DM-MMN. Fix an arbitrary $T \subseteq \mathcal{I}$ such that $T^c \cap \mathcal{D} \neq \emptyset$, and choose a node $d \in T^c \cap \mathcal{D}$. Fix

an arbitrary $\lambda \in (1, \infty)$. Let $s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ be a λ -simulating distribution of $p_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ neglecting T such that $s_{W_{\mathcal{I}}, X_{\mathcal{I}}^n, Y_{\mathcal{I}}^n, \hat{W}_{\mathcal{I} \times \mathcal{I}}}$ satisfies all the properties in Lemma 4. Then, it follows from Proposition 1 and Definition 2 with the identifications $U \equiv W_T$, $V \equiv \hat{W}_{T \times \{d\}}$, $p_{U,V} \equiv p_{W_T, \hat{W}_{T \times \{d\}}}$, $q_V \equiv s_{\hat{W}_{T \times \{d\}}}$, $|\mathcal{W}| \equiv 2^{\sum_{i \in T} \lceil nR_i \rceil}$ and $\alpha \equiv \Pr\{W_T \neq \hat{W}_{T \times \{d\}}\} \leq \bar{\epsilon}$ that

$$\begin{aligned} & D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \\ & \geq \sum_{i \in T} nR_i + \lambda(\lambda - 1)^{-1} \log(1 - \alpha) \\ & \geq \sum_{i \in T} nR_i + \lambda(\lambda - 1)^{-1} \log(1 - \bar{\epsilon}). \end{aligned} \quad (25)$$

B. Using the DPI to Introduce the Channel Input and Output

Let $U_{T^c}^{k-1} \triangleq (W_{\mathcal{I}}, X_{\mathcal{I}}^{k-1}, Y_{T^c}^{k-1})$ and $V_{T^c}^{k-1} \triangleq (W_{T^c}, X_{T^c}^{k-1}, Y_{T^c}^{k-1})$ be the random variables generated before the k^{th} time slot, and consider the following chain of inequalities:

$$\begin{aligned} & D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \\ & \stackrel{(a)}{\leq} D_\lambda(p_{U_{T^c}^0, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{V_{T^c}^0, \hat{W}_{T \times \{d\}}}) \\ & = D_\lambda(p_{U_{T^c}^0, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{V_{T^c}^0} s_{\hat{W}_{T \times \{d\}} | V_{T^c}^0}) \\ & \stackrel{(b)}{=} D_\lambda(p_{U_{T^c}^0, \hat{W}_{T \times \{d\}}} \| p_{W_T} p_{V_{T^c}^0} s_{\hat{W}_{T \times \{d\}} | V_{T^c}^0}) \\ & \stackrel{(c)}{=} D_\lambda(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | U_{T^c}^0} \| p_{U_{T^c}^0} s_{\hat{W}_{T \times \{d\}} | V_{T^c}^0}) \\ & \stackrel{(d)}{\leq} D_\lambda(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}}, Y_{T^c}^n | U_{T^c}^0} \| p_{U_{T^c}^0} s_{\hat{W}_{T \times \{d\}}, Y_{T^c}^n | V_{T^c}^0}) \\ & = D_\lambda(p_{U_{T^c}^0} p_{Y_{T^c}^n | U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | U_{T^c}^0, Y_{T^c}^n} \| p_{U_{T^c}^0} s_{Y_{T^c}^n | V_{T^c}^0} s_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n}) \\ & \stackrel{(e)}{=} D_\lambda(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} p_{Y_{T^c}^n | U_{T^c}^0} \| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} s_{Y_{T^c}^n | V_{T^c}^0}) \\ & \stackrel{(f)}{\leq} D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} p_{X_{\mathcal{I}}^n, Y_{T^c}^n | U_{T^c}^0} \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} s_{X_{\mathcal{I}}^n, Y_{T^c}^n | V_{T^c}^0} \prod_{k=1}^n p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}} \right\| \right) \end{aligned} \quad (26)$$

where

- (a) follows from the DPI of D_λ by introducing W_{T^c} .
- (b) follows from Property (i) in Lemma 4.
- (c) follows from the fact that W_T and W_{T^c} are independent.
- (d) follows from the DPI of D_λ by introducing the channel output $Y_{T^c}^n$.
- (e) follows from Property (ii) in Lemma 4 and the fact that

$$(W_T \rightarrow (W_{T^c}, Y_{T^c}^n) \rightarrow \hat{W}_{T \times \{d\}})_p$$

forms a Markov chain.

- (f) follows from the DPI of D_λ by introducing the channel input $X_{\mathcal{I}}^n$.

In order to simplify (26), we consider

$$\begin{aligned}
& D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} p_{X_{\mathcal{I}}^n | U_{T^c}^0} \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} s_{X_{T^c}^n | V_{T^c}^0} \prod_{k=1}^n p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}} \right\| \right) \\
&= D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n p_{X_{\mathcal{I},k} | Y_{T^c,k} | U_{T^c}^{k-1}} \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (s_{X_{T^c,k} | Y_{T^c,k} | U_{T^c}^{k-1}} p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}}) \right\| \right) \\
&\stackrel{(1)}{=} D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} q_{Y_{T^c} | X_{\mathcal{I}}}) \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (s_{X_{T^c,k} | V_{T^c}^{k-1}} s_{Y_{T^c,k} | V_{T^c}^{k-1}, X_{T^c,k}} p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}}) \right\| \right) \\
&\stackrel{(a)}{=} D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} q_{Y_{T^c} | X_{\mathcal{I}}}) \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (s_{X_{T^c,k} | V_{T^c}^{k-1}} s_{Y_{T^c,k} | X_{T^c,k}} p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}}) \right\| \right) \\
&\stackrel{(b)}{=} D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} q_{Y_{T^c} | X_{\mathcal{I}}}) \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{T^c,k} | U_{T^c}^{k-1}} p_{X_{T,k} | U_{T^c}^{k-1}, X_{T^c,k}} s_{Y_{T^c,k} | X_{T^c,k}}) \right\| \right) \\
&= D_\lambda \left(p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} q_{Y_{T^c} | X_{\mathcal{I}}}) \left\| p_{U_{T^c}^0} p_{\hat{W}_{T \times \{d\}} | V_{T^c}^0, Y_{T^c}^n} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} s_{Y_{T^c,k} | X_{T^c,k}}) \right\| \right), \tag{27}
\end{aligned}$$

where

(a) follows from Property (iii) in Lemma 4.

(b) follows from Property (v) in Lemma 4.

C. Single-Letterizing the Rényi Divergence

Consider the distribution

$$\begin{aligned}
p_{U_{T^c}^0, X_{\mathcal{I}}^n, Y_{T^c}^n} &= p_{U_{T^c}^0} p_{X_{\mathcal{I}}^n | Y_{T^c}^n | U_{T^c}^0} \\
&= p_{U_{T^c}^0} \prod_{k=1}^n p_{X_{\mathcal{I},k} | Y_{T^c,k} | U_{T^c}^{k-1}} \\
&= p_{U_{T^c}^0} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} p_{Y_{T^c,k} | U_{T^c}^{k-1}, X_{\mathcal{I},k}}) \\
&\stackrel{(1)}{=} p_{U_{T^c}^0} \prod_{k=1}^n (p_{X_{\mathcal{I},k} | U_{T^c}^{k-1}} q_{Y_{T^c} | X_{\mathcal{I}}}). \tag{28}
\end{aligned}$$

Using (26), (27) and Definition 7 and omitting subscripts of probability distributions to simplify notation, we have

$$\begin{aligned}
& D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \\
& \leq \frac{1}{\lambda - 1} \log \left(\sum_{u_{T^c}^0, x_{\mathcal{I}}^n, y_{T^c}^n, \hat{w}_{T \times \{d\}}} p(u_{T^c}^0) p(\hat{w}_{T \times \{d\}} | v_{T^c}^0, y_{T^c}^n) \prod_{k=1}^n \left(p(x_{\mathcal{I},k} | u_{T^c}^{k-1}) q(y_{T^c,k} | x_{\mathcal{I},k}) \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \right) \right) \\
& \stackrel{(28)}{=} \frac{1}{\lambda - 1} \log \left(\sum_{u_{T^c}^0, x_{\mathcal{I}}^n, y_{T^c}^n, \hat{w}_{T \times \{d\}}} p(u_{T^c}^0, x_{\mathcal{I}}^n, y_{T^c}^n) p(\hat{w}_{T \times \{d\}} | v_{T^c}^0, y_{T^c}^n) \prod_{k=1}^n \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \right) \\
& = \frac{1}{\lambda - 1} \log \left(\sum_{x_{\mathcal{I}}^n, y_{T^c}^n} p(x_{\mathcal{I}}^n, y_{T^c}^n) \prod_{k=1}^n \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \right). \tag{29}
\end{aligned}$$

Following (29), we consider the following chain of equalities:

$$\begin{aligned}
& \sum_{x_{\mathcal{I}}^n, y_{T^c}^n} p(x_{\mathcal{I}}^n, y_{T^c}^n) \prod_{k=1}^n \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \\
& = \sum_{x_{\mathcal{I}}^n, y_{T^c}^n} \prod_{k=1}^n \left(p(x_{\mathcal{I},k}, y_{T^c,k} | x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}) \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \right) \\
& \stackrel{(1)}{=} \sum_{x_{\mathcal{I}}^n, y_{T^c}^n} \prod_{k=1}^n \left(p(x_{\mathcal{I},k} | x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}) \frac{(q(y_{T^c,k} | x_{\mathcal{I},k}))^\lambda}{(s(y_{T^c,k} | x_{T^c,k}))^{\lambda-1}} \right). \tag{30}
\end{aligned}$$

Letting $f_0^{(\lambda, T)}(x_{\mathcal{I}}^0, y_{T^c}^0) \triangleq 1$ and

$$f_k^{(\lambda, T)}(x_{\mathcal{I}}^k, y_{T^c}^k) \triangleq \prod_{\ell=1}^k \left(p(x_{\mathcal{I},\ell} | x_{\mathcal{I}}^{\ell-1}, y_{T^c}^{\ell-1}) \frac{(q(y_{T^c,\ell} | x_{\mathcal{I},\ell}))^\lambda}{(s(y_{T^c,\ell} | x_{T^c,\ell}))^{\lambda-1}} \right) \tag{31}$$

for each $k \in \{1, 2, \dots, n\}$ and following (30), we consider

$$\begin{aligned}
& \log \sum_{x_{\mathcal{I}}^n, y_{T^c}^n} \prod_{k=1}^n \left(p(x_{\mathcal{I},k} | x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}) \frac{(q(y_{T^c,k} | x_{\mathcal{I},k}))^\lambda}{(s(y_{T^c,k} | x_{T^c,k}))^{\lambda-1}} \right) \\
& \stackrel{(31)}{=} \log \sum_{x_{\mathcal{I}}^n, y_{T^c}^n} f_n^{(\lambda, T)}(x_{\mathcal{I}}^n, y_{T^c}^n) \\
& \stackrel{(a)}{=} \log \prod_{k=1}^n \frac{\sum_{x_{\mathcal{I}}^k, y_{T^c}^k} f_k^{(\lambda, T)}(x_{\mathcal{I}}^k, y_{T^c}^k)}{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} f_{k-1}^{(\lambda, T)}(x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1})} \\
& = \sum_{k=1}^n \log \left(\frac{\sum_{x_{\mathcal{I}}^k, y_{T^c}^k} f_k^{(\lambda, T)}(x_{\mathcal{I}}^k, y_{T^c}^k)}{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} f_{k-1}^{(\lambda, T)}(x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1})} \right), \tag{32}
\end{aligned}$$

where (a) is a telescoping product. For each $k \in \{1, 2, \dots, n\}$, define $p_{X_{\mathcal{I},k}}^{(\lambda, T)}$ to be the following distribution:

$$p_{X_{\mathcal{I},k}}^{(\lambda, T)}(x_{\mathcal{I},k}) \triangleq \frac{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} p(x_{\mathcal{I},k} | x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}) f_{k-1}^{(\lambda, T)}(x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1})}{\sum_{x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1}} f_{k-1}^{(\lambda, T)}(x_{\mathcal{I}}^{k-1}, y_{T^c}^{k-1})} \tag{33}$$

for all $x_{\mathcal{I},k}$. Combining (30), (32) and (33), we obtain

$$\log \left(\sum_{x_{\mathcal{I}}^n, y_{T^c}^n} p(x_{\mathcal{I}}^n, y_{T^c}^n) \prod_{k=1}^n \left(\frac{q(y_{T^c,k} | x_{\mathcal{I},k})}{s(y_{T^c,k} | x_{T^c,k})} \right)^{\lambda-1} \right) = \sum_{k=1}^n \log \left(\sum_{x_{\mathcal{I},k}} p_{X_{\mathcal{I},k}}^{(\lambda,T)}(x_{\mathcal{I},k}) \sum_{y_{T^c,k}} \frac{(q(y_{T^c,k} | x_{\mathcal{I},k}))^\lambda}{(s(y_{T^c,k} | x_{T^c,k}))^{\lambda-1}} \right),$$

which implies from (29) and Definition 7 that

$$D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \leq \sum_{k=1}^n D_\lambda(q_{Y_{T^c} | X_{\mathcal{I}}} \| s_{Y_{T^c,k} | X_{T^c,k}} p_{X_{\mathcal{I},k}}^{(\lambda,T)}). \quad (34)$$

D. Representing Distributions in the Rényi Divergence by a Single Distribution

Construct a probability distribution $p_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda,T)}$ for each $k \in \{1, 2, \dots, n\}$ as

$$p_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda,T)}(x_{\mathcal{I}}, y_{T^c}) \triangleq p_{X_{\mathcal{I},k}}^{(\lambda,T)}(x_{\mathcal{I}}) q_{Y_{T^c} | X_{\mathcal{I}}}(y_{T^c} | x_{\mathcal{I}}) \quad (35)$$

for all $(x_{\mathcal{I}}, y_{T^c})$ (cf. (33)), where $q_{Y_{T^c} | X_{\mathcal{I}}}$ denotes the channel of the DM-MMN. Combining (31), (33), (35) and Property (iv) in Lemma 4, we have

$$p_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda,T)} = s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda,T)} \quad (36)$$

for each $k \in \{1, 2, \dots, n\}$ where $s_{X_{\mathcal{I},k}, Y_{T^c,k}}^{(\lambda,T)}$ is as defined in (14). Then, it follows from Property (iv) in Lemma 4, (36) and (34) that

$$D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \leq \sum_{k=1}^n D_\lambda(q_{Y_{T^c} | X_{\mathcal{I}}} \| p_{Y_{T^c,k} | X_{T^c,k}}^{(\lambda,T)} p_{X_{\mathcal{I},k}}^{(\lambda,T)}). \quad (37)$$

Using (35) and Proposition 3, we obtain

$$p_{Y_{T^c,k} | X_{\mathcal{I},k}}^{(\lambda,T)} = q_{Y_{T^c} | X_{\mathcal{I}}} \quad (38)$$

for all $k \in \{1, 2, \dots, n\}$, which implies from (37) that

$$D_\lambda(p_{W_T, \hat{W}_{T \times \{d\}}} \| p_{W_T} s_{\hat{W}_{T \times \{d\}}}) \leq \sum_{k=1}^n D_\lambda(p_{Y_{T^c,k} | X_{\mathcal{I},k}}^{(\lambda,T)} \| p_{Y_{T^c,k} | X_{T^c,k}}^{(\lambda,T)} p_{X_{\mathcal{I},k}}^{(\lambda,T)}). \quad (39)$$

E. Introduction of a Time-sharing Random Variable

Let Q_n be a random variable uniformly distributed on $\{1, 2, \dots, n\}$ and independent of all other random variables. Construct the probability distribution $p_{Q_n, X_{\mathcal{I}}^n, Y_{T^c}^n}^{(\lambda,T)}$ such that

$$p_{Q_n, X_{\mathcal{I}}^n, Y_{T^c}^n}^{(\lambda,T)}(k, x_{\mathcal{I}}^n, y_{T^c}^n) = \frac{1}{n} \prod_{h=1}^n p_{X_{\mathcal{I},h}, Y_{T^c,h}}^{(\lambda,T)}(x_{\mathcal{I},h}, y_{T^c,h}) \quad (40)$$

for all $k \in \{1, 2, \dots, n\}$, $x_{\mathcal{I}}^n \in \mathcal{X}_{\mathcal{I}}^n$ and $y_{T^c}^n \in \mathcal{Y}_{T^c}^n$. Then, we can calculate the joint distributions $p_{Q_n, X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n}^{(\lambda, T)}$ and $p_{X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n}^{(\lambda, T)}$ as follows:

$$\begin{aligned} p_{Q_n, X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n}^{(\lambda, T)}(k, x_{\mathcal{I}}, y_{T^c}) &= p_{Q_n}^{(\lambda, T)}(k) p_{X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n | Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c} | k) \\ &= p_{Q_n}^{(\lambda, T)}(k) p_{X_{\mathcal{I}, k}, Y_{T^c}, k | Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c} | k) \\ &\stackrel{(40)}{=} \frac{1}{n} p_{X_{\mathcal{I}, k}, Y_{T^c}, k}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c}) \end{aligned} \quad (41)$$

and

$$\begin{aligned} p_{X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c}) &\stackrel{(41)}{=} \frac{1}{n} \sum_{k=1}^n p_{X_{\mathcal{I}, k}, Y_{T^c}, k}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c}) \\ &\stackrel{(38)}{=} \frac{1}{n} \sum_{k=1}^n p_{X_{\mathcal{I}, k}}^{(\lambda, T)}(x_{\mathcal{I}}) q_{Y_{T^c} | X_{\mathcal{I}}}(y_{T^c} | x_{\mathcal{I}}) \\ &\stackrel{(41)}{=} p_{X_{\mathcal{I}, Q_n}}^{(\lambda, T)}(x_{\mathcal{I}}) q_{Y_{T^c} | X_{\mathcal{I}}}(y_{T^c} | x_{\mathcal{I}}). \end{aligned} \quad (42)$$

It follows from (41) and Proposition 3 that

$$p_{X_{\mathcal{I}, Q_n}, Y_{T^c}, Q_n | Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c} | k) = p_{X_{\mathcal{I}, k}, Y_{T^c}, k}^{(\lambda, T)}(x_{\mathcal{I}}, y_{T^c}), \quad (43)$$

$$(Q_n \rightarrow X_{\mathcal{I}, Q_n} \rightarrow Y_{T^c}, Q_n)_{s(\lambda, T)} \quad (44)$$

and

$$(Q_n \rightarrow X_{T^c, Q_n} \rightarrow Y_{T^c}, Q_n)_{s(\lambda, T)}. \quad (45)$$

Following (39), consider the following chain of inequalities:

$$\begin{aligned} &\frac{1}{n} \sum_{k=1}^n D_{\lambda}(p_{Y_{T^c}, k | X_{\mathcal{I}, k}}^{(\lambda, T)} \| p_{Y_{T^c}, k | X_{T^c}, k}^{(\lambda, T)} | p_{X_{\mathcal{I}, k}}^{(\lambda, T)}) \\ &= \frac{1}{(\lambda - 1)n} \sum_{k=1}^n \log \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}, k}}^{(\lambda, T)}(x_{\mathcal{I}}) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, k | X_{\mathcal{I}, k}}^{(\lambda, T)}(y_{T^c} | x_{\mathcal{I}}))^{\lambda}}{(p_{Y_{T^c}, k | X_{T^c}, k}^{(\lambda, T)}(y_{T^c} | x_{T^c}))^{\lambda-1}} \\ &\stackrel{(43)}{=} \frac{1}{(\lambda - 1)n} \sum_{k=1}^n \log \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}, Q_n} | Q_n}^{(\lambda, T)}(x_{\mathcal{I}} | k) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, Q_n | X_{\mathcal{I}, Q_n}, Q_n}^{(\lambda, T)}(y_{T^c} | x_{\mathcal{I}}, k))^{\lambda}}{(p_{Y_{T^c}, Q_n | X_{T^c}, Q_n, Q_n}^{(\lambda, T)}(y_{T^c} | x_{T^c}, k))^{\lambda-1}} \\ &\stackrel{(a)}{\leq} \frac{1}{(\lambda - 1)} \log \sum_{k=1}^n \frac{1}{n} \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}, Q_n} | Q_n}^{(\lambda, T)}(x_{\mathcal{I}} | k) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, Q_n | X_{\mathcal{I}, Q_n}, Q_n}^{(\lambda, T)}(y_{T^c} | x_{\mathcal{I}}, k))^{\lambda}}{(p_{Y_{T^c}, Q_n | X_{T^c}, Q_n, Q_n}^{(\lambda, T)}(y_{T^c} | x_{T^c}, k))^{\lambda-1}} \\ &\stackrel{(b)}{=} \frac{1}{(\lambda - 1)} \log \sum_{k=1}^n \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}, Q_n}, Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, k) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, Q_n | X_{\mathcal{I}, Q_n}, Q_n}^{(\lambda, T)}(y_{T^c} | x_{\mathcal{I}}, k))^{\lambda}}{(p_{Y_{T^c}, Q_n | X_{T^c}, Q_n, Q_n}^{(\lambda, T)}(y_{T^c} | x_{T^c}, k))^{\lambda-1}} \\ &\stackrel{(c)}{=} \frac{1}{(\lambda - 1)} \log \sum_{k=1}^n \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}, Q_n}, Q_n}^{(\lambda, T)}(x_{\mathcal{I}}, k) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, Q_n | X_{\mathcal{I}, Q_n}}^{(\lambda, T)}(y_{T^c} | x_{\mathcal{I}}))^{\lambda}}{(p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda, T)}(y_{T^c} | x_{T^c}))^{\lambda-1}} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{(\lambda - 1)} \log \sum_{x_{\mathcal{I}}} p_{X_{\mathcal{I}}, Q_n}^{(\lambda, T)}(x_{\mathcal{I}}) \sum_{y_{T^c}} \frac{(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}(y_{T^c} | x_{\mathcal{I}}))^{\lambda}}{(p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}(y_{T^c} | x_{T^c}))^{\lambda-1}} \\
&= D_{\lambda}(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}^{(\lambda, T)} \| p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda, T)} | p_{X_{\mathcal{I}}, Q_n}^{(\lambda, T)}), \tag{46}
\end{aligned}$$

where

- (a) follows from the concavity of $t \mapsto \log t$ and Jensen's inequality.
- (b) follows from (40) that $p_{Q_n}^{(\lambda, T)}(k) = 1/n$ for all $k \in \{1, 2, \dots, n\}$.
- (c) follows from (44) and (45).

Combining (25), (39) and (46), we obtain

$$\sum_{i \in T} n R_i + \lambda(\lambda - 1)^{-1} \log(1 - \bar{\epsilon}) \leq n D_{\lambda}(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}^{(\lambda, T)} \| p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda, T)} | p_{X_{\mathcal{I}}, Q_n}^{(\lambda, T)}) \tag{47}$$

for all $\lambda \in (1, \infty)$.

F. Approximating the Rényi Divergence by Conditional Relative Entropy

For each block length n , choose λ to be dependent on n as follows:

$$\lambda_n \triangleq 1 + \frac{1}{\sqrt{n}}.$$

It then follows from (47), Proposition 2, and the fact that $|\mathcal{X}_T| |\mathcal{Y}_{T^c}| \leq |\mathcal{X}_{\mathcal{I}}| |\mathcal{Y}_{\mathcal{I}}|$ that

$$\begin{aligned}
\sum_{i \in T} R_i + \left(\frac{1}{n} + \frac{1}{\sqrt{n}} \right) \log(1 - \bar{\epsilon}) &\leq D_{\lambda_n}(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)} \| p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda_n, T)} | p_{X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)}) \\
&\leq D(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)} \| p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda_n, T)} | p_{X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)}) + \frac{8(|\mathcal{X}_{\mathcal{I}}| |\mathcal{Y}_{\mathcal{I}}|)^5}{\sqrt{n}} \tag{48}
\end{aligned}$$

if $n \geq 16$ (i.e., $\lambda_n \leq 5/4$ so Proposition 2 applies). Taking the limit inferior on both sides of (48), we obtain

$$\sum_{i \in T} R_i \leq \liminf_{n \rightarrow \infty} D(p_{Y_{T^c}, Q_n | X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)} \| p_{Y_{T^c}, Q_n | X_{T^c}, Q_n}^{(\lambda_n, T)} | p_{X_{\mathcal{I}}, Q_n}^{(\lambda_n, T)}). \tag{49}$$

Consider each distribution on $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{T^c})$ as a point in the $|\mathcal{X}_{\mathcal{I}}| |\mathcal{Y}_{T^c}|$ -dimensional Euclidean space. Then, by the compactness of the probability simplex, there exists a subsequence of the natural numbers $\{1, 2, \dots\}$, say indexed by $\{n_{\ell}\}_{\ell=1}^{\infty}$, such that $\{p_{X_{\mathcal{I}}, Q_{n_{\ell}}}, Y_{T^c}, Q_{n_{\ell}}}\}_{\ell=1}^{\infty}$ is convergent with respect to the \mathcal{L}_1 -distance. Let $\bar{p}_{X_{\mathcal{I}}, Y_{T^c}}$ be the limit of the subsequence such that

$$\bar{p}_{X_{\mathcal{I}}, Y_{T^c}}(x_{\mathcal{I}}, y_{T^c}) = \lim_{\ell \rightarrow \infty} p_{X_{\mathcal{I}}, Q_{n_{\ell}}, Y_{T^c}, Q_{n_{\ell}}}^{(\lambda_{n_{\ell}}, T)}(x_{\mathcal{I}}, y_{T^c}) \tag{50}$$

for all $(x_{\mathcal{I}}, y_{T^c})$. Combining (42) and (50), we have

$$\bar{p}_{X_{\mathcal{I}}, Y_{T^c}}(x_{\mathcal{I}}, y_{T^c}) = \bar{p}_{X_{\mathcal{I}}}(x_{\mathcal{I}}) q_{Y_{T^c} | X_{\mathcal{I}}}(y_{T^c} | x_{\mathcal{I}}). \tag{51}$$

Since $D(p_{Y_{T^c}|X_{\mathcal{I}}} \| p_{Y_{T^c}|X_{T^c}} | p_{X_{\mathcal{I}}})$ is a continuous functional of distribution $p_{X_{\mathcal{I}}, Y_{T^c}}$, it follows from (49) and (50) that

$$\begin{aligned}
\sum_{i \in T} R_i &\leq D(\bar{p}_{Y_{T^c}|X_{\mathcal{I}}} \| \bar{p}_{Y_{T^c}|X_{T^c}} | \bar{p}_{X_{\mathcal{I}}}) \\
&= \sum_{x_{\mathcal{I}}} \bar{p}_{X_{\mathcal{I}}}(x_{\mathcal{I}}) \sum_{y_{T^c}} \bar{p}_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c}|x_{\mathcal{I}}) \log \frac{\bar{p}_{Y_{T^c}|X_{\mathcal{I}}}(y_{T^c}|x_{\mathcal{I}})}{\bar{p}_{Y_{T^c}|X_{T^c}}(y_{T^c}|x_{T^c})} \\
&= \sum_{x_{\mathcal{I}}, y_{T^c}} \bar{p}_{X_{\mathcal{I}}, Y_{T^c}}(x_{\mathcal{I}}, y_{T^c}) \log \frac{\bar{p}_{X_{\mathcal{I}}, Y_{T^c}|X_{T^c}}(x_{\mathcal{I}}, y_{T^c}|x_{T^c})}{\bar{p}_{X_{\mathcal{I}}|X_{T^c}}(x_{\mathcal{I}}|x_{T^c}) \bar{p}_{Y_{T^c}|X_{T^c}}(y_{T^c}|x_{T^c})} \\
&= I_{\bar{p}_{X_{\mathcal{I}}, Y_{T^c}}}(X_{\mathcal{I}}; Y_{T^c} | X_{T^c}). \tag{52}
\end{aligned}$$

The theorem then follows from (51) and (52).

G. Comparison to the Proof of the Cut-Set Bound using Fano's Inequality

Following the setting in Section VI-A at the beginning of the proof of Theorem 1 and following the cut-set bound approach that uses Fano's inequality [1, Theorem 18.1] (leading to a weak converse), we can lower bound the average error probability ϵ_n as follows

$$(1 - \epsilon_n) \sum_{i \in T} n R_i \leq I_{p_{W_{\mathcal{I}}, Y_{T^c}^n}}(W_T; Y_{T^c}^n | W_{T^c}). \tag{53}$$

The bound (53) holds for each T that satisfies $T^c \cap \mathcal{D} \neq \emptyset$. Next, using the DPI for the relative entropy and a time-sharing random variable for the purpose of single-letterization [1, Theorem 18.1], it can be shown that

$$I_{p_{W_{\mathcal{I}}, Y_{T^c}^n}}(W_T; Y_{T^c}^n | W_{T^c}) \leq n I_{\bar{p}_{X_{\mathcal{I}}, Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \tag{54}$$

where $\bar{p}_{X_{\mathcal{I}}}(x_{\mathcal{I}}) = \frac{1}{n} \sum_{k=1}^n p_{X_{\mathcal{I}}, k}(x_{\mathcal{I}})$ is the empirical input distribution induced by the $(n, R_{\mathcal{I}})$ -code. Combining (53) and (54) and using the fact that $\bar{p}_{X_{\mathcal{I}}}(x_{\mathcal{I}})$ does not depend on T , we obtain

$$\mathcal{C}_{\epsilon} \subseteq \bigcup_{\bar{p}_{X_{\mathcal{I}}}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \frac{1}{1-\epsilon} I_{\bar{p}_{X_{\mathcal{I}}, Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \tag{55}$$

For $\epsilon = 0$, (55) immediately reduces to the cut-set bound (5). For $\epsilon > 0$, the bound in (55) cannot be used to prove strong converse theorems because of the multiplicative factor $\frac{1}{1-\epsilon}$.

The proofs of (3) and (55) share many common steps, but significantly they differ in the first step where for fixed rates, lower bounds on the error probabilities are sought. More specifically, our approach relates a conditional Rényi divergence to the error probability (cf. Proposition 1), while the approach that hinges on Fano's inequality relates a conditional mutual information to the error probability (cf. the inequality in (53)). However beyond the first step, the application of the DPI and the method of single-letterization are almost the same for both proofs, but we do need to eventually approximate the conditional Rényi entropy with the conditional mutual information (cf. Proposition 2) to obtain bound (3). The two different ways of lower bounding the error probability yield two

different outer bounds stated in (3) and (55) respectively.

VII. CLASSES OF MULTIMESSAGE MULTICAST NETWORKS WITH TIGHT CUT-SET BOUND

In this section, we will use Theorem 1 to prove strong converses for some classes of DM-MMNs whose capacity regions are known. Unless specified otherwise, we let (S, \mathcal{D}) denote the multicast demand on the networks.

A. Multicast Networks with Maximal Cut-Set Distribution

We start this section by stating an achievability result for multimessage multicast networks in the following theorem, which is a specialization of the main result of *noisy network coding* by Lim, Kim, El Gamal and Chung [25]. Noisy network coding was also discovered by Yassaee and Aref [26].

Theorem 3: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a DM-MMN, and let

$$\mathcal{R}_{\text{in}} \triangleq \bigcup_{\prod_{i=1}^N p_{X_i}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) - H_{p_{X_{\mathcal{I}}} q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}}(Y_T | X_{\mathcal{I}}, Y_{T^c}) \\ \text{where } p_{X_{\mathcal{I}}} \triangleq \prod_{i=1}^N p_{X_i}, \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (56)$$

Then, $\mathcal{R}_{\text{in}} \subseteq \mathcal{C}_0$.

Proof: The theorem follows by taking $\hat{Y} = Y$ in Theorem 1 of [25]. ■

We would like to identify multicast networks whose inner bounds \mathcal{R}_{in} coincides with our outer bound \mathcal{R}_{out} in Theorem 1. Using the following definition and corollary, we can state, in Theorem 4, a sufficient condition for $\mathcal{R}_{\text{in}} = \mathcal{R}_{\text{out}}$ to hold.

Definition 8: A DM-MMN $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ is said to be *dominated by a maximal product distribution* if there exists some product distribution $p_{X_{\mathcal{I}}}^* \triangleq \prod_{i=1}^N p_{X_i}^*$ such that the following statement holds for each $T \subseteq \mathcal{I}$:

$$I_{p_{X_{\mathcal{I}}}^* q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) = \max_{p_{X_{\mathcal{I}}}} \{ I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) \}.$$

The following corollary is a direct consequence of Theorem 1 and Definition 8, and the proof is deferred to Appendix B.

Corollary 5: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a DM-MMN, and let

$$\mathcal{R}_{\text{out}}^* \triangleq \bigcup_{\prod_{i=1}^N p_{X_i}} \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{(\prod_{i=1}^N p_{X_i}) q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (57)$$

If the DM-MMN is dominated by a maximal product distribution, then $\mathcal{C}_{\epsilon} \subseteq \mathcal{R}_{\text{out}}^*$ for all $\epsilon \in [0, 1)$.

Theorem 4: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a DM-MMN. Suppose the DM-MMN satisfies the following two conditions:

1. The DM-MMN is dominated by a maximal product distribution.
2. For all $T \subseteq \mathcal{I}$ and all $p_{X_{\mathcal{I}}}$, $H_{p_{X_{\mathcal{I}}} q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}}(Y_T | X_{\mathcal{I}}, Y_{T^c}) = 0$.

Then $\mathcal{R}_{\text{in}} = \mathcal{C}_{\epsilon} = \mathcal{R}_{\text{out}}^*$ for all $\epsilon \in [0, 1)$.

Proof: Since the DM-MMN is dominated by a maximal product distribution, it follows from Theorem 3 and Corollary 5 that $\mathcal{R}_{\text{in}} \subseteq \mathcal{C}_0 \subseteq \mathcal{C}_\epsilon \subseteq \mathcal{R}_{\text{out}}^*$ for all $\epsilon \in [0, 1]$. In addition, it follows from (56), (57) and Condition 2 that $\mathcal{R}_{\text{in}} = \mathcal{R}_{\text{out}}^*$. ■

Theorem 4 implies the strong converse for the classes of DM-MMNs which satisfy Conditions 1 and 2. Since the deterministic relay networks with no interference [6], the finite-field linear deterministic networks [7], [8] and the wireless erasure networks [9] satisfy both conditions in Theorem 4, the strong converse holds for these networks. We note that for the class of wireless erasure networks, one assumes that the erasure pattern of the entire network is known to each destination, i.e., Y_d contains the erasure pattern as side information for each $d \in \mathcal{D}$ [9, Section III.C], and hence Condition 2 in Theorem 4 is satisfied. In the following subsection, we introduce a DM-MMN connected by independent DMCs and prove the strong converse using Corollary 5 and Theorem 1.

B. DM-MMN Consisting of Independent DMCs

Consider a DM-MMN where a DMC is defined for every link $(i, j) \in \mathcal{I} \times \mathcal{I}$. Let $\mathcal{X}_{i,j}$ and $\mathcal{Y}_{i,j}$ denote the input and output alphabets of the DMC carrying information from node i to node j for each $(i, j) \in \mathcal{I} \times \mathcal{I}$, and let $q_{Y_{i,j}|X_{i,j}}$ denote the DMC. For each $(i, j) \in \mathcal{I} \times \mathcal{I}$, the capacity of channel $q_{Y_{i,j}|X_{i,j}}$, denoted by $C_{i,j}$, is attained by some $\bar{p}_{X_{i,j}}$, i.e.,

$$\begin{aligned} C_{i,j} &\triangleq \max_{p_{X_{i,j}}} I_{p_{X_{i,j}} q_{Y_{i,j}|X_{i,j}}} (X_{i,j}; Y_{i,j}) \\ &= I_{\bar{p}_{X_{i,j}} q_{Y_{i,j}|X_{i,j}}} (X_{i,j}; Y_{i,j}). \end{aligned} \quad (58)$$

Then, we define the input and output alphabets for each node i in the following natural way:

$$\mathcal{X}_i \triangleq \mathcal{X}_{i,1} \times \mathcal{X}_{i,2} \times \dots \times \mathcal{X}_{i,N}$$

and

$$\mathcal{Y}_i \triangleq \mathcal{Y}_{1,i} \times \mathcal{Y}_{2,i} \times \dots \times \mathcal{Y}_{N,i} \quad (59)$$

for each $i \in \mathcal{I}$, and we let $q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}$ denote the channel of the network. In addition, we assume

$$q_{Y_{\mathcal{I}}|X_{\mathcal{I}}} = \prod_{(i,j) \in \mathcal{I} \times \mathcal{I}} q_{Y_{i,j}|X_{i,j}}, \quad (60)$$

i.e., the random transformations (noises) from $X_{i,j}$ to $Y_{i,j}$ are independent and the overall channel of the network is in a product form. It then follows from (60) and Proposition 3 that

$$((\{X_{k,\ell}\}_{(k,\ell) \neq (i,j)}, \{Y_{k,\ell}\}_{(k,\ell) \neq (i,j)}) \rightarrow X_{i,j} \rightarrow Y_{i,j})_{p_{X_{\mathcal{I}}} q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}} \quad (61)$$

forms a Markov chain for all $(i, j) \in \mathcal{I} \times \mathcal{I}$. We call the network described above the *DM-MMN consisting of independent DMCs*. One important example of such networks is the *line network* in which $\mathcal{I} \times \mathcal{I}$ consists of nonzero-capacity links of the form $(i, i+1)$ for all $i \in \{1, 2, \dots, N-1\}$ and zero-capacity links for the other node

pairs. Define

$$\mathcal{R}' \triangleq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq \sum_{(i,j) \in T \times T^c} C_{i,j}, \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\}. \quad (62)$$

Since the DMCs from $X_{i,j}$ to $Y_{i,j}$ are all independent and each of the DMC can carry information at a rate arbitrarily close to the capacity, it follows from the network equivalence theory [27] and Theorem 3 that \mathcal{R}' is precisely the capacity region of the DM-MMN consisting of independent DMCs, which is formally stated in the following corollary and proved in Appendix C.

Corollary 6: $\mathcal{R}' = \mathcal{C}_0$.

We use the outer bound \mathcal{R}_{out} proved in Theorem 1 (cf. (2)) to prove the following lemma.

Lemma 7: $\mathcal{C}_\epsilon \subseteq \mathcal{R}_{\text{out}} \subseteq \mathcal{R}'$ for all $\epsilon \in [0, 1)$.

For completeness, the proof is provided in Appendix D. The following theorem is a direct consequence of Corollary 6 and Lemma 7.

Theorem 5: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be a DM-MMN consisting of independent DMCs. Then, $\mathcal{C}_\epsilon = \mathcal{R}'$ for all $\epsilon \in [0, 1)$.

Theorem 5 implies the strong converse for the class of DM-MMNs consisting of independent DMCs.

C. Single-Destination DM-MMN Consisting of Independent DMCs with Destination Feedback

In this section, we examine a class of DM-MMNs with destination feedback, which is a generalization of the DM-MMN consisting of independent DMCs discussed in the previous section. We assume $|\mathcal{D}| = 1$ and let $d \in \mathcal{I}$ denote the (single) destination node throughout this section. We define the single-destination DM-MMN consisting of independent DMCs with feedback as follows.

Definition 9: Let $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ be DM-MMN consisting of independent DMCs with multicast demand $(\mathcal{S}, \{d\})$ as defined in the previous section. A single-destination DM-MMN with multicast demand $(\mathcal{S}, \{d\})$, denoted by $(\mathcal{X}_{\mathcal{I}}, \tilde{\mathcal{Y}}_{\mathcal{I}}, \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}})$, is called the *feedback version of* $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ if the following two conditions hold:

- 1) $\tilde{\mathcal{Y}}_i = \mathcal{Y}_i \times \mathcal{Y}_d$ for all $i \in \mathcal{I}$.
- 2) Suppose $(X_{\mathcal{I}}, Y_{\mathcal{I}}) \in \mathcal{X}_{\mathcal{I}} \times \mathcal{Y}_{\mathcal{I}}$ associated with the MMN $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ is generated according to $p_{X_{\mathcal{I}}} q_{Y_{\mathcal{I}}|X_{\mathcal{I}}}$ for some input distribution $p_{X_{\mathcal{I}}}$. Then, the random tuple $(\tilde{X}_{\mathcal{I}}, \tilde{Y}_{\mathcal{I}}) \in \mathcal{X}_{\mathcal{I}} \times \tilde{\mathcal{Y}}_{\mathcal{I}}$ associated with the MMN $(\mathcal{X}_{\mathcal{I}}, \tilde{\mathcal{Y}}_{\mathcal{I}}, \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}})$ is distributed according to $p_{X_{\mathcal{I}}} \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}}$ where $\tilde{X}_i = X_i$ and $\tilde{Y}_i = (Y_i, Y_d)$ for all $i \in \mathcal{I}$.

Let $(\mathcal{X}_{\mathcal{I}}, \tilde{\mathcal{Y}}_{\mathcal{I}}, \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}})$ be the feedback version of $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ with multicast demand $(\mathcal{S}, \{d\})$. It then follows from Definitions 9 and 2 that for any $(n, R_{\mathcal{I}})$ -code on $(\mathcal{X}_{\mathcal{I}}, \tilde{\mathcal{Y}}_{\mathcal{I}}, \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}})$, both Y_d^{k-1} and Y_i^{k-1} are available for encoding $X_{i,k}$ at node i for all $i \in \mathcal{I}$. In other words, there exists for each $i \in \mathcal{I}$ a perfect feedback link which carries the output symbols at node d to node i . Consequently, the capacity region of $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ is always a subset of the capacity region of $(\mathcal{X}_{\mathcal{I}}, \tilde{\mathcal{Y}}_{\mathcal{I}}, \tilde{q}_{\tilde{Y}_{\mathcal{I}}|X_{\mathcal{I}}})$. Shannon showed in [28] that the capacity of any DMC is equal to the capacity of the feedback version, and the strong converse for the feedback version has been shown in [11, Section IV]. Also see [29, Problem 2.5.16(c)] for another proof sketch of the strong converse for a DMC with

feedback. Here, we show that \mathcal{R}' (defined in (62)) is equal to the ϵ -capacity region of any single-destination DM-MMN consisting of independent DMCs as well as the ϵ -capacity region of the feedback version for any $\epsilon \in [0, 1)$. In other words, feedback does not enlarge the ϵ -capacity region of any single-destination DM-MMN consisting of independent DMCs. Thus, the strong converse also holds for the feedback version of this class of DM-MMNs.

Theorem 6: Let $(\mathcal{X}_I, \mathcal{Y}_I, q_{Y_I|X_I})$ be a DM-MMN consisting of independent DMCs with multicast demand $(\mathcal{S}, \{d\})$, and let \mathcal{R}' be the set defined in (62). Suppose $(\mathcal{X}_I, \tilde{\mathcal{Y}}_I, \tilde{q}_{\tilde{Y}_I|X_I})$ is a feedback version of $(\mathcal{X}_I, \mathcal{Y}_I, q_{Y_I|X_I})$. Let $\epsilon \in [0, 1)$ be a real number and let \mathcal{C}_ϵ and $\tilde{\mathcal{C}}_\epsilon$ be the ϵ -capacity regions of $(\mathcal{X}_I, \mathcal{Y}_I, q_{Y_I|X_I})$ and $(\mathcal{X}_I, \tilde{\mathcal{Y}}_I, \tilde{q}_{\tilde{Y}_I|X_I})$ respectively. Then, $\tilde{\mathcal{C}}_\epsilon = \mathcal{C}_\epsilon = \mathcal{R}'$.

Theorem 6 can be proved similarly to Theorem 5. We provide a concise proof in Appendix E. Since the ϵ -capacity region with *imperfect feedback* compared with perfect feedback cannot be larger and the ϵ -capacity region with *no feedback* is equal to \mathcal{R}' by Theorem 5, it follows from Theorem 6 that the strong converse also holds for any single-destination DM-MMN consisting of independent DMCs with imperfect feedback.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we proved that the strong converse holds for some classes of DM-MMNs for which the cut-set bound is achievable by leveraging some elementary properties of the conditional Rényi divergence. We suggest three promising avenues for future research. First, the foremost item is to show that all rate tuples that lie in the exterior of the usual cut-set bound for DM-MMNs [2] result in error probabilities tending to one. This seems rather challenging as we have to assert the existence of a *common* distribution \bar{p}_{X_I, Y_I} for all cut-sets T in (52). This would allow us to swap the intersection and union in Theorem 1. Second, and less ambitiously, we also hope to extend our result to Gaussian networks [1, Chapter 19], which may be tractable if we restrict the models under consideration to the class of Gaussian networks for which the optimum input distribution is a multivariate Gaussian. Finally, it may be fruitful and instructive to focus our attention on smaller DM-MMNs such as the DM-RC.

APPENDIX A

PROOF OF PROPOSITION 2

Proof of Proposition 2: For any random variables U and V , we let

$$\mathcal{S}_{U|v} \triangleq \{u \in \mathcal{U} : \mathbf{Pr}\{U = u | V = v\} > 0\}$$

be the *support* of U conditioned on the event $\{V = v\}$. If V is a trivial random variable, i.e., $\mathcal{V} = \emptyset$, then \mathcal{S}_U is simply the support of U . If $\lambda = 1$, the statement of the proposition is obvious so henceforth, we prove the statement for $\lambda \in (1, 5/4]$. Suppose (X, Y, Z) is jointly distributed according to $p_{X,Y,Z}(x, y, z)$ which we abbreviate as $p(x, y, z)$ in this proof. Let

$$g(\lambda) \triangleq \log \sum_{z \in \mathcal{S}_Z} p(z) \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x, y|z))^\lambda}{(p(x|z)p(y|z))^{\lambda-1}}$$

be a function of λ defined on $[1, \infty)$. Straightforward calculations involving l'Hôpital's rule reveal that $g(1) = 0$ and $g'(1) = D(p_{X,Y|Z} || p_{X|Z} p_{Y|Z} | p_Z)$ (cf. Definition 7). Using Taylor's theorem, we obtain

$$g(\lambda) = g(1) + (\lambda - 1)g'(1) + (\lambda - 1)^2 \frac{g''(a)}{2}$$

for some $a \in [1, \lambda]$, which implies that

$$g(\lambda) = (\lambda - 1)D(p_{X,Y|Z} || p_{X|Z} p_{Y|Z} | p_Z) + (\lambda - 1)^2 \frac{g''(a)}{2}. \quad (63)$$

Using standard calculus techniques, we obtain

$$\begin{aligned} g''(a) = & \frac{\sum_{z \in \mathcal{S}_Z} p(z) \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \left(\log \frac{p(x,y|z)}{p(x|z)p(y|z)} \right)^2}{\sum_{z \in \mathcal{S}_Z} p(z) \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}}} \\ & - \left(\frac{\sum_{z \in \mathcal{S}_Z} p(z) \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \log \frac{p(x,y|z)}{p(x|z)p(y|z)}}{\sum_{z \in \mathcal{S}_Z} p(z) \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}}} \right)^2. \end{aligned} \quad (64)$$

In order to obtain an upper bound for $|g''(a)|$, we will calculate a lower bound for

$$\sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}}$$

and upper bounds for

$$\sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \log \frac{p(x,y|z)}{p(x|z)p(y|z)}$$

and

$$\sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \left(\log \frac{p(x,y|z)}{p(x|z)p(y|z)} \right)^2.$$

Consider the following chain of inequalities:

$$\begin{aligned} \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} \frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} & \geq \sum_{(x,y) \in \mathcal{S}_{X,Y|z}} (p(x,y|z))^a \\ & \geq \max_{x,y} (p(x,y|z))^a \\ & \geq (|\mathcal{X}||\mathcal{Y}|)^{-a} \\ & \geq (|\mathcal{X}||\mathcal{Y}|)^{-5/4}. \end{aligned} \quad (65)$$

On the other hand, fix x, y and z such that $p(z) > 0$ and $p(x, y|z) > 0$, and consider $\frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \log \frac{p(x,y|z)}{p(x|z)p(y|z)}$ as well as $\frac{(p(x,y|z))^a}{(p(x|z)p(y|z))^{a-1}} \left(\log \frac{p(x,y|z)}{p(x|z)p(y|z)} \right)^2$. Since $\min\{p(x|z), p(y|z)\} \geq p(x, y|z)$, there exist $0 \leq k_1 \leq 1$ and $0 \leq k_2 \leq 1$ such that $p(x|z) = (p(x, y|z))^{k_1}$ and $p(y|z) = (p(x, y|z))^{k_2}$. Using the facts that $a \in (1, 5/4]$ and

$0 \leq k_1 + k_2 \leq 2$, we have

$$|1 - (k_1 + k_2)| \leq 1 \quad (66)$$

and

$$a - (a - 1)(k_1 + k_2) \geq 1/2 \quad (67)$$

Then,

$$\begin{aligned} \frac{(p(x, y|z))^a}{(p(x|z)p(y|z))^{a-1}} \log \frac{p(x, y|z)}{p(x|z)p(y|z)} &= (p(x, y|z))^{a-(a-1)(k_1+k_2)} \log(p(x, y|z))^{1-(k_1+k_2)} \\ &\stackrel{(a)}{\leq} (p(x, y|z))^{1/2} \log(p(x, y|z))^{-1} \\ &= 2(p(x, y|z))^{1/2} \log(p(x, y|z))^{-1/2} \\ &\stackrel{(b)}{\leq} 2e^{-1} \log e \end{aligned} \quad (68)$$

and

$$\begin{aligned} \frac{(p(x, y|z))^a}{(p(x|z)p(y|z))^{a-1}} \left(\log \frac{p(x, y|z)}{p(x|z)p(y|z)} \right)^2 &= (p(x, y|z))^{a-(a-1)(k_1+k_2)} (\log(p(x, y|z))^{1-(k_1+k_2)})^2 \\ &\stackrel{(c)}{\leq} (p(x, y|z))^{1/2} (\log(p(x, y|z))^{-1})^2 \\ &= 4(p(x, y|z))^{1/2} (\log(p(x, y|z))^{-1/2})^2 \\ &\stackrel{(d)}{\leq} 16e^{-2} \log e, \end{aligned} \quad (69)$$

where

- (a) follows from (66) and (67);
- (b) follows from calculus that $q \log q^{-1} \leq e^{-1} \log e$ for all $0 < q \leq 1$;
- (c) follows from (66) and (67);
- (d) follows from calculus that $q(\log q^{-1})^2 \leq 4e^{-2} \log e$ for all $0 < q \leq 1$.

Combining (63), (64), (65), (68) and (69), we obtain

$$g(\lambda) \leq (\lambda - 1)D(p_{X,Y|Z}||p_{X|Z}p_{Y|Z}|p_Z) + 8(\lambda - 1)^2(|\mathcal{X}||\mathcal{Y}|)^5,$$

which implies that for each $\lambda \in (1, 5/4]$ (note $\lambda \neq 1$ so we can cancel the common factors $\lambda - 1$),

$$D_\lambda(p_{X,Y|Z}||p_{X|Z}p_{Y|Z}|p_Z) \leq D(p_{X,Y|Z}||p_{X|Z}p_{Y|Z}|p_Z) + 8(\lambda - 1)(|\mathcal{X}||\mathcal{Y}|)^5$$

and hence (9) follows. ■

APPENDIX B

PROOF OF COROLLARY 5

Proof of Corollary 5: Suppose the DM-MMN is dominated by some maximal product distribution $p_{X_{\mathcal{I}}}^* \triangleq \prod_{i=1}^N p_{X_i}^*$ such that for each $T \subseteq \mathcal{I}$, we have

$$I_{p_{X_{\mathcal{I}}}^* q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) = \max_{p_{X_{\mathcal{I}}}} \{I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c})\}.$$

This then implies from Theorem 1 that for each $\epsilon \in [0, 1)$,

$$\mathcal{C}_{\epsilon} \subseteq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \mathcal{D} \neq \emptyset} \left\{ R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}}}^* q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right. \right\} \stackrel{(57)}{\subseteq} \mathcal{R}_{\text{out}}^*.$$

This completes the proof. \blacksquare

APPENDIX C

PROOF OF COROLLARY 6

Proof of Corollary 6: Construct a counterpart of the channel $(\mathcal{X}_{\mathcal{I}}, \mathcal{Y}_{\mathcal{I}}, q_{Y_{\mathcal{I}}|X_{\mathcal{I}}})$ as follows: Let $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{X}}_{\mathcal{I}}, \bar{q}_{\bar{\mathcal{X}}_{\mathcal{I}}|\bar{\mathcal{X}}_{\mathcal{I}}})$ be a noiseless DM-MMN consisting of independent DMCs with multicast demand $(\mathcal{S}, \mathcal{D})$ such that for each $(i, j) \in \mathcal{I} \times \mathcal{I}$, the DMC carrying information from node i to node j is an error-free (noiseless) channel, denoted by $\bar{q}_{\bar{\mathcal{X}}_{i,j}|\bar{\mathcal{X}}_{i,j}}$, with capacity $C_{i,j}$ (cf. (58)). To be more precise, $\bar{q}_{\bar{\mathcal{X}}_{i,j}|\bar{\mathcal{X}}_{i,j}}$ can carry $\lfloor nC_{i,j} \rfloor$ error-free bits for each $(i, j) \in \mathcal{I} \times \mathcal{I}$ for n uses of $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{X}}_{\mathcal{I}}, \bar{q}_{\bar{\mathcal{X}}_{\mathcal{I}}|\bar{\mathcal{X}}_{\mathcal{I}}})$. Let $\bar{\mathcal{C}}$ denote the capacity region of $(\bar{\mathcal{X}}_{\mathcal{I}}, \bar{\mathcal{X}}_{\mathcal{I}}, \bar{q}_{\bar{\mathcal{X}}_{\mathcal{I}}|\bar{\mathcal{X}}_{\mathcal{I}}})$. It follows from the network equivalence theory [27] that $\mathcal{C}_0 = \bar{\mathcal{C}}$. In addition, it has been shown in [25, Section IIA] that $\bar{\mathcal{C}} = \mathcal{R}'$. Consequently, $\mathcal{R}' = \bar{\mathcal{C}} = \mathcal{C}_0$, which is what was to be proved. \blacksquare

APPENDIX D

PROOF OF LEMMA 7

Proof of Lemma 7: Since $\mathcal{C}_{\epsilon} \subseteq \mathcal{R}_{\text{out}}$ for all $\epsilon \in [0, 1)$ by Theorem 1, it remains to show $\mathcal{R}_{\text{out}} \subseteq \mathcal{R}'$. In order to obtain an outer bound of \mathcal{R}_{out} , we consider the following chain of inequalities for each $p_{X_{\mathcal{I}}}$ and each $T \subseteq \mathcal{I}$:

$$\begin{aligned} & I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}) \\ &= \sum_{j \in T^c} I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_j | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}) \\ &\stackrel{(59)}{=} \sum_{j \in T^c} \sum_{\ell=1}^N I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{\ell,j} | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\ &\stackrel{(a)}{=} \sum_{j \in T^c} \sum_{\ell \in T} I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_T; Y_{\ell,j} | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\ &\stackrel{(b)}{=} \sum_{j \in T^c} \sum_{\ell \in T} I_{p_{X_{\mathcal{I}}} q_{Y_{T^c}|X_{\mathcal{I}}}}(X_{\ell,j}; Y_{\ell,j}) \\ &\stackrel{(58)}{\leq} \sum_{(\ell,j) \in T \times T^c} C_{\ell,j}. \end{aligned} \tag{70}$$

where

(a) follows from the fact that for all $\ell \in T^c$,

$$\begin{aligned}
& I_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; Y_{\ell,j} | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\
&= H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) - H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{\mathcal{I}}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\
&\stackrel{(61)}{=} H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{\ell,j}) - H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{\ell,j}) \\
&= 0;
\end{aligned}$$

(b) follows from the fact that for all $\ell \in T$,

$$\begin{aligned}
& I_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; Y_{\ell,j} | X_{T^c}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\
&\leq H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j}) - H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{\mathcal{I}}, Y_{\{\bar{j} \in T^c: \bar{j} < j\}}, \{Y_{m,j}\}_{m=1}^{\ell-1}) \\
&\stackrel{(61)}{=} H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j}) - H_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(Y_{\ell,j} | X_{\ell,j}) \\
&= I(X_{\ell,j}; Y_{\ell,j}).
\end{aligned}$$

Combining (2), (62) and (70), we have $\mathcal{R}_{\text{out}} \subseteq \mathcal{R}'$. ■

APPENDIX E

PROOF OF THEOREM 6

Proof of Theorem 6: Fix any $\epsilon \in [0, 1)$. Since $\mathcal{C}_\epsilon = \mathcal{R}'$ by Theorem 5 and $\mathcal{C}_\epsilon \subseteq \tilde{\mathcal{C}}_\epsilon$, it remains to show that $\tilde{\mathcal{C}}_\epsilon \subseteq \mathcal{R}'$. Define

$$\tilde{\mathcal{R}}_{\text{out}} \triangleq \bigcap_{T \subseteq \mathcal{I}: T^c \cap \{d\} \neq \emptyset} \bigcup_{R_{\mathcal{I}} \left| \begin{array}{l} \sum_{i \in T} R_i \leq I_{p_{X_{\mathcal{I}} \tilde{q}_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; \tilde{Y}_{T^c} | X_{T^c}), \\ R_i = 0 \text{ for all } i \in \mathcal{S}^c \end{array} \right.} \left\{ R_{\mathcal{I}} \right\}. \quad (71)$$

Since $\tilde{\mathcal{C}}_\epsilon \subseteq \tilde{\mathcal{R}}_{\text{out}}$ for all $\epsilon \in [0, 1)$ by Theorem 1 and $\mathcal{R}_{\text{out}} \subseteq \mathcal{R}'$ by Lemma 7, it suffices to show $\tilde{\mathcal{R}}_{\text{out}} = \mathcal{R}_{\text{out}}$. To this end, we consider the following chain of equalities for each $p_{X_{\mathcal{I}}}$ and each $T \subseteq \mathcal{I}$ such that $T^c \cap \{d\} \neq \emptyset$:

$$\begin{aligned}
I_{p_{X_{\mathcal{I}} \tilde{q}_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; \tilde{Y}_{T^c} | X_{T^c}) &\stackrel{(a)}{=} I_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; Y_{T^c}, Y_d | X_{T^c}) \\
&\stackrel{(b)}{=} I_{p_{X_{\mathcal{I}} q_{Y_{T^c}} | X_{\mathcal{I}}}}(X_T; Y_{T^c} | X_{T^c}).
\end{aligned} \quad (72)$$

where

(a) follows from Condition 2 in Definition 9;

(b) follows from the fact that $T^c \cap \{d\} \neq \emptyset$.

Combining (2), (71) and (72), we have $\mathcal{R}_{\text{out}} = \tilde{\mathcal{R}}_{\text{out}}$. ■

ACKNOWLEDGMENTS

The authors are indebted to Prof. Shun Watanabe for pointing out an error in an earlier version of this paper.

REFERENCES

- [1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.
- [2] A. El Gamal, "On information flow in relay networks," in *Proc. IEEE National Telecom Conference*, vol. 2, Nov 1981, pp. D4.1.1–D4.1.4.
- [3] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Inf. Theory*, vol. 25, no. 5, pp. 572–84, 1979.
- [4] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channel," *IEEE Trans. on Inf. Theory*, vol. 55, no. 3, pp. 921 – 930, 2009.
- [5] A. E. Gamal and M. R. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. on Inf. Theory*, vol. 28, no. 3, p. 536, 1982.
- [6] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2425–2432, 2006.
- [7] A. S. Avestimehr, S. N. Diggavi, and D. N. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.
- [8] E. Perron, "Information-theoretic secrecy for wireless networks," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
- [9] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, 2006.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.
- [11] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. Allerton Conference on Communication, Control and Computing*, Sep. 2010, pp. 1327 – 1333.
- [12] D. Xu and D. Erdogmuns, "Rényi's entropy, divergence and their nonparametric estimators," in *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*, J. C. Principe, Ed. Springer, 2010, pp. 47–102.
- [13] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, pp. 26–34, Jan. 1995.
- [14] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. on Inf. Theory*, vol. 19, no. 5, pp. 357–359, 1973.
- [15] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Trans. on Inf. Theory*, vol. 45, no. 7, pp. 2486–2489, 1999.
- [16] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Commun. Math. Phys.*, vol. 331, no. 2, pp. 593–622, 2014.
- [17] V. Y. F. Tan, "On the reliability function of the discrete memoryless relay channel," *IEEE Trans. on Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, 2015.
- [18] A. Behboodi and P. Piantanida, "On the asymptotic error probability of composite relay channels," in *Proc. of IEEE Intl. Symp. on Inf. Theory*, St Petersburg, Russia, 2011, pp. 1524 – 1528.
- [19] —, "On the asymptotic spectrum of the error probability of composite networks," in *Proc. of IEEE Inf. Theory Workshop*, Lausanne, Switzerland, 2012, pp. 148–152.
- [20] A. Behboodi, "Cooperative networks with channel uncertainty," Ph.D. dissertation, Department of Telecommunications, Supélec (École Supérieure d'Électricité), 2012.
- [21] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2003.
- [22] T. van Erven and P. Harremoës, "Rényi divergence and Kullback–Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.
- [24] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2008.
- [25] S. H. Lim, Y.-H. Kim, A. El Gamal and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, 2011.
- [26] M. Yassaee and M. R. Aref, "Slepian-Wolf coding over cooperative relay networks," *IEEE Trans. on Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, Jun. 2011.
- [27] R. Köetter, M. Effros, and M. Médard, "A theory of network equivalence — Part I: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, 2011.

- [28] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, 1956.
- [29] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.